DOCUMENT RESUME

ED 250 373 TM 840 688

AUTHOR Adams, Raymond J.
TITLE Sex Bias in ASAT? ACER Research Monograph No. 24.
INSTITUTION Australian Council for Educational Research, Hawthorn.
REPORT NO ISBN-0-85563-376-X
PUB DATE 84
NOTE 164p.
AVAILABLE FROM Australian Council for Educational Research Limited, Radford House, Frederick Street, Hawthorn, Victoria 3122.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Academic Persistence; College Entrance Examinations; Course Content; Foreign Countries; *National Competency Tests; School Registration; Secondary Education; *Sex Differences; Student Attitudes; *Test Bias; Test Coaching; Test Items; *Test Results
IDENTIFIERS Australia; *Australian Scholastic Aptitude Tests

ABSTRACT

The Australian Scholastic Test (ASAT) was first used as a moderating device in the Australian Capitol Territory in 1977. Differences have been observed in the average performance of males and females on the test. The study reported in this monograph investigated the nature and origin of the score differences recorded for ASAT The research focused on five issues: retention rate differences, attitude differences, preparation differences, item bias, and effect of differential course selection. The findings indicated that sex had no significant direct effect on ASAT scores. The observed differences in male and female ASAT scores were related to English ability, experience in mathematics, and confidence in success. The different patterns for staying longer at school could explain much of the observed variations between males and females. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ACER Research Monograph No. 24

ED250373

Sex Bias in ASAT?

Raymond J. Adams

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

J. Williams

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 810 088

ACER RESEARCH MONOGRAPH NO. 24

SEX BIAS IN ASAT?

RAYMOND J. ADAMS

Australian Council for Educational Research

Hawthorn, Victoria.

1984

Published by
The Australian Council for Educational Research Limited
Radford House, Frederick Street, Hawthorn, Victoria 3122

Printed and bound by
Brown Prior Anderson Pty Ltd
5 Evans Street, Burwood, Victoria 3125

Cover design by Louise Coutts

National Library Cataloguing-in-Publication data.

Adams, Raymond J, 1959 -.
Sex bias in ASAT?

Bibliography
ISBN 0 85563 376 X.

1. Test bias - Australia. 2. Educational tests and measurements - Australia. 3. Students, Rating of - Australia. 4. Sex discrimination in education. I. Australian Council for Educational Research. II. Title. (Series: ACER research monograph; no. 24).

371.2'6'013

This research study was supported by grants from:
The Commonwealth Department of Education and Youth Affairs,
The ACT Schools Authority,
The Board of Secondary School Studies, Queensland,
The Tertiary Institutions Service Centre, Western Australia.

Copyright © ACER 1984

TABLE OF CONTENTS

		Page
LIST OF TABLES		vii
LIST OF FIGURES		viii
ACKNOWLEDGMENTS		ix
CHAPTER 1	SEX DIFFERENCES AND ASAT	1
	Scholastic Aptitude Tests in Australia	1
	ASAT and Its Use	1
	The Problem	2
	The ACT Score Adjustment	3
	The Problem in Perspective	4
	Mathematics Achievement in the ACT	5
	Problems in Sex Differences Research	5
	Outcomes of the Study	6
CHAPTER 2	A REVIEW OF PREVIOUS RESEARCH	7
	Defining Bias	7
	Differences in Abilities	7
	Differences in Preparation	9
	Differences in Course Type	11
	Differences in Apprehension and Approach to Aptitude Tests	14
	Item Bias	15
	Differences in Retentivity	17
	Conclusion	18
CHAPTER 3	OUTLINE OF THE STUDY	19
	Some Questions for Investigation	19
	Examining the Data	21
	The Key Issues	22
	Aims of the Study	23
CHAPTER 4	RETENTION AND SCORE DIFFERENCES	24
	Alternative Models	25
	Summary and Conclusions	39
CHAPTER 5	INTERVIEWS WITH STUDENTS	41
	Australian Capital Territory	42
	Queensland	50
	Summary and Conclusions	55
CHAPTER 6	STUDENT ATTITUDES TOWARDS ASAT	59
	Scale Development	60
	The Samples	61
	Factor Analyses	62
	Results	64
	Analysis of Variance	67
	Discussion	70
	Summary and Conclusions	72
CHAPTER 7	TIME SPENT IN PREPARATION	74
	Results	75
	Overview	78

		Page
CHAPTER 8	ITEM BIAS IN ASAT-L	79
	Results	81
	Discussion	83
	Summary and Conclusions	88
CHAPTER 9	DIFFERENT FACTORS AND THEIR EFFECTS	90
	A First Model	91
	A Second Model	95
	Accounting for the Observed Difference	98
	Summary and Conclusions	99
CHAPTER 10	THE FINDINGS AND THEIR IMPLICATIONS	101
	Results of the Study	102
	Implications of the Study and Its Findings	107
	Conclusions	109
BIBLIOGRAPHY		111
APPENDIXES		
Appendix 4.1	Notes on the Calculation of the Sizes of the Age Cohorts	119
Appendix 4.2	The Truncated Normal Distribution	120
Appendix 5.1	ASAT General Information Interview Schedule	124
Appendix 6.1	ASAT General Information Questionnaire	132
Appendix 6.2	Factor Analysis of Items	142
Appendix 8.1	Traditional Item Statistics for ASAT-L	143
Appendix 8.2	Results of Factor Analyses of Items and Varimax Rotation with Two Factors	147
Appendix 8.3	Results of Factor Analyses of Units and Varimax Rotation with Two Factors	150
Appendix 8.4	Rasch Analyses Item Statistics for ASAT-L	151

LIST OF TABLES

		Page
Table 3.1	Mean Raw Scores on ASAT and Its Subtests	20
Table 4.1	Proportion of Age Cohort Requiring Tertiary Entrance Score at Year 12	25
Table 4.2	Student Mean Scores on ASAT	25
Table 4.3	Proportion of Male Age Cohort Retained with CGS Included	27
Table 4.4	Comparison of the Difference between Male and Female Mean Scores and Difference between Male and Modified Female Mean Scores	30
Table 4.5	Difference between Male and Modified Female Mean Scores with the Inclusion of Canberra Grammar School	31
Table 4.6	Expected and Observed Differences in ASAT Scores	34
Table 4.7	Correlation between Observed and Expected Differences in Scores for the Modified Model	38
Table 5.1	Use of Complex Strategies by ACT Students in Answering ASAT	47
Table 6.1	Size and Sex Composition of Samples	62
Table 6.2	Factor Patterns of Attitude Scale Items	63
Table 6.3	Reliabilities for Attitude Scales	64
Table 6.4	Attitude Scale Statistics for each State	65
Table 6.5	Analysis of Variance Results for Confidence in Success on ASAT	68
Table 6.6	Analysis of Variance Results for Effectance Motivation	69
Table 6.7	Analysis of Variance Results for ASAT as a Male Domain	69
Table 6.8	Analysis of Variance Results for Attitude Towards Success on ASAT	70
Table 7.1	Analysis of Variance Results for School Preparation Sessions	76
Table 7.2	Analysis of Variance Results for Home Preparation Time	76
Table 7.3	Analysis of Variance Results for School Preparation Sessions with Respect to School Type	77
Table 7.4	Analysis of Variance Results for Home Preparation Time with Respect to School Type	78
Table 8.1	Intercorrelations between the Subscales	82
Table 8.2	Summary of Possible Elements of Bias	84

Page

Table 9.1 Results of Regression Analyses with Confidence in Success Included 92

Table 9.2 Results of Regression Analyses with Effectance Motivation Included 92

Table 9.3 Results of Regression Analyses with Attitude Towards Success Included 92

Table 9.4 Results of Regression Analyses with Moderated English 97

Table 9.5 Results of Regression Analyses with Unmoderated English 97

Table 9.6 Means of Variables in the Model 98

LIST OF FIGURES

Page

Figure 4.1 Plot of Retention Rate Differences and ASAT Score Differences 26

Figure 4.2 Plot of Retention Rate Differences and ASAT Score Differences with CGS Included 28

Figure 4.3 A Hypothetical Distribution of ASAT Scores for the Age Cohort, showing the proportion q, above the cut-off point, k 33

Figure 4.4 Plot of the Expected Differences in ASAT Score Against the Observed Differences in ASAT Score 35

Figure 4.5 Correlation Surface Relating ASAT Score to the Retentivity Variable 36

Figure 4.6 Plot of the Expected Difference Against the Observed Difference, $\rho_m = 0.8$, $\rho_f = 0.9$ 39

Figure 6.1 Profiles of Student Scores 64

Figure 6.2 Comparison of Male and Female Scores 65

Figure 9.1 A Proposed Causal Model 91

Figure 9.2 Causal Model with Confidence in Success Included 95

Figure 9.3 A Causal Model including English 96

Figure 9.4 Causal Model with Unmoderated English 99

Figure 9.5 Contribution of Variables to the Observed Sex Difference 100

ACKNOWLEDGMENTS

A report of this nature cannot be undertaken without the assistance of many people. I would firstly like to thank all of my colleagues at the ACER for their encouragement and constructive criticism.

Acknowledgments must also go to Mrs Dinah Rawlinson and Mrs Merle Warry for their assistence with interviewing students, to Dr Gilah Leder for her help in developing the questionnaire, to Dr Darryl Daly for his suggestions in relation to the retention rate material and Professor Hedley Beare who provided valuable help and advice. An acknowledgment must also go to the examining authorities of the Australian Capital Territory, Queensland and Western Australia, without whose cooperation the study would not have been possible.

I must thank Mrs Yvonne Allen who typed the final manuscript with skill and patience while under considerable pressure and my wife, Kathi, who spent many hours reading draft manuscripts and providing important advice. When not reading she was providing invaluable support and encouragement. In addition, Mr M. Plunkett and the staff at the Swinburne Computer Centre made possible the analyses undertaken in this report; for their help I am very grateful.

CHAPTER 1

SEX DIFFERENCES AND ASAT

Scholastic Aptitude Tests in Australia

The problems associated with the selection of students from secondary level to tertiary level education have given rise to many surveys, research projects, reports and literature reviews.

Several selection methods are currently being employed or have been used in the past in different parts of Australia ranging from external examinations to school based assessment. Variations and combinations of both of these methods are currently being used.

Until the early 1970s, public examinations without any internal assessment component were used in all Australian States. However, in no State did these tests have a high predictive power for university performance. In 1967, the University of Western Australia, the Australian National University and the Department of Education in Tasmania made a submission to the Commonwealth Department of Education and Science for funds to explore the possible use of an aptitude test battery in the selection of students for tertiary education. In 1968, the Australian Council for Educational Research (ACER) was commissioned to develop an aptitude test battery suitable for use in a projected study of tertiary selection. The project was known as the Tertiary Education Entrance Project (TEEP) and the first battery of tests (Series A) was produced later that year. The tests were successively modified and in 1970 the first of the Australian Scholastic Aptitude Tests (ASAT) was developed. Currently a new series of ASAT is developed each year for use in the Australian Capital Territory, Queensland and Western Australia.

ASAT and Its Use

In its present form ASAT consists of 100 multiple choice items based on a variety of material drawn from the Humanities, Social Sciences, Science and Mathematics. These subjects are represented in the following proportions:

30 per cent Humanities
20 per cent Social Science
30 per cent Science
20 per cent Mathematics

After ASAT is administered it is divided into verbal and quantitative subscales.

The questions are designed so that it will be possible to answer them even if a student is not familiar with the area from which the questions are drawn. ASAT is also designed to be independent of any school curriculum, undertaken by students through the selection of courses. It aims to test a student's capacity for intellectual pursuits at the tertiary level, as well as aptitude to undertake studies at the Year 12 level.

The basic use made of ASAT is to moderate between different subjects and schools in order to bring students' marks onto a common scale. In Western Australia, ASAT is used to moderate between subjects that are assessed by an external examination. In the Australian Capital Territory and Queensland where the students' performance is assessed within the school, ASAT is used to moderate between schools and subjects. In both these systems the group average ASAT scores and standard deviations are used to adjust students' scores before calculating Tertiary Entrance Scores. The details for the moderation procedure for the ACT Schools Authority are provided in a paper by Keeves, McBryde and Bennett (1978) and for Queensland in an article by McGaw (1977).

In Western Australia, 10 per cent of a student's ASAT score is added to the tertiary entrance score. In Queensland and the Australian Capital Territory a student's ASAT does not contribute directly to the tertiary entrance score, but in the Australian Capital Territory the score has appeared on the student's certificate, and has been taken into consideration in the selection procedures employed at the Australian National University.

The Problem

Since 1977, when ASAT was first used in the Australian Capital Territory as a moderating device, a difference has been observed between males and females in their performance on this test. These differences have also been evident in each of the two other systems that use ASAT, but to a lesser extent. These differences have usually resulted from males outscoring the females on the quantitative subscale and the females outscoring the males, to a lesser degree, on the verbal subscale. Although the sex difference in mean ASAT scores has existed since 1977, not until 1981 and 1982 were the differences regarded as substantial. In the ACT in 1981, the male mean ASAT score exceeded the female mean ASAT score by 5.1 points, and in 1982, the difference was 6.8 points.

Since sex differences were also being observed in Queensland and Western Australia, representatives from the three systems resolved in June 1983 to undertake a research project to investigate the possible bias in ASAT. It was hoped that the study would look at all forms of possible bias, not just gender bias.

The ACT Score Adjustment

When the large sex difference between ASAT scores occurred in 1982, the ACT Schools Authority came under pressure from the single sex girls schools within the ACT to do something about the so called 'sex bias'. At about the same time Senator Ryan, the Commonwealth Minister for Education and Youth Affairs, expressed the view that the differences between male and female scores were unacceptable.

In 1983, the ACT Schools Accrediting Agency, a standing committee of the ACT Schools Authority, recognized the impracticability of gathering any significant additional data before making a decision for that year. It therefore established an independent Review Panel to consider the existing data and advise on the procedures to be employed in 1983. The report of this Review Panel was presented on 19th September, 1983.

Noting that it was unable to conduct a thorough review in the time available the ASAT Review Panel made the following points.

1 A significant proportion of the difference in boys' and girls' scores could be explained by the existence of a greater proportion of females in the system. This was based on evidence available to the panel that indicated that up to 50 per cent of the sex difference might be attributed to differences in retention.

2 The remaining difference was believed to be due to such factors as curriculum choice and probable bias in ASAT.

3 The ACT Schools Accrediting Agency should move to reduce the test bias for 1983. There would be insufficient time in that year to rewrite and retrial the test, and thus the Agency should apply a post-test correction factor.

4 After considering options the Review Panel believed the correction factor should be one third of the difference between the means of the standardized ASAT scores of female and male students.

The ACT Schools Authority accepted these recommendations as an interim measure for 1983 alone. This resulted in the female students being allocated an extra 0.995 score points on their 1983 ASAT scores. These adjusted scores were used in the calculation of the standardized scores for performance on ASAT recorded on the students' certificates and in the application of moderation procedures between schools and subjects for the calculation of moderated subject and course scores and the Tertiary Entrance Score.

The investigation presented in this report is concerned with the obtaining of evidence that might assist in the resolution of the issue of whether or not sex bias existed in the Australian Scholastic Aptitude Test and whether, if such bias of a systematic kind existed, it could be corrected for by an adjustment to the scores of one group of students.

The Problem in Perspective

Over the last century, perhaps more studies have been undertaken in the fields of education, psychology and sociology in relation to sex differences than on any other single topic. This research has been accompanied by substantial changes in societal attitudes towards the education of females. Surveys by Anastasi (1958), Maccoby (1966) and Maccoby and Jacklin (1974) reported that United States boys had superior spatial and numerical ability while the females usually did better in verbal and linguistic areas. It was however, emphasized that the sex differences were small when compared to within sex standard deviations.

Historically, women were believed to have developed differently in the course of evolution. It was believed that they were less able in the areas of abstract reasoning, and in general had physical and mental incapacities. One of the first objections to this view came in 1906 from E.L. Thorndike. He argued that the differences between males and females were too small to be of practical significance and they probably resulted from social influences rather than biological differences. Views of this type have given support to the feminist movement and as a result, interest has been generated in a search for evidence of the equality of the sexes.

In the 1920s and 1930s research focused on attempts to differentiate accurately between the characteristics of the same sex group based on measures of masculinity and femininity. Many studies were related to psychoanalytic theory which hypothesized that emotional differences arose from biological rather than sociological sources. In the 1950s, the focus again shifted to

socialization and the study of sex role effects. The research extended to cover the investigation of influences associated with biological, psychological, social and cognitive development. A review of the more recent research has been carried out and is provided in Chapter 2. The reported research indicated that in general, sex differences are small when compared to within sex differences and are largely unpredictable. Any male advantage would appear to be in quantitative and spatial abilities, while the female advantage would appear to be in the areas associated with language and verbal skills. The current views on the origins of these differences lie in differences in socialization for males and females, which lead to differences between the sexes in career expectations, in expectations of success in mathematics and verbal studies and in attitudes and values towards mathematics and language studies. A likely cause of the sex difference in performance on ASAT could be associated with sex differences in achievement and participation in mathematics programs at the Year 12 level.

Mathematics Achievement in the ACT

Evidence presented by Moss (1982) showed that in marked contrast to the findings recorded for other State educational systems in Australia, girls at the lower secondary level in the Australian Capital Territory performed in 1978 at a significantly higher level than did boys on a mathematics test including basic arithmetic, algebra and lower mental process subtests. On the geometry and higher mental process subtests, the girls also recorded a higher level of performance than the boys did, but the difference was not statistically significant.

At the Year 12 level in 1978, a substantially higher proportion of girls (38.6 per cent of the age cohort) in the ACT continued with the study of mathematics, while in Queensland the proportion was 17.6 per cent, and in Western Australia 18.2 per cent. Thus in the ACT compared to other parts of Australia, partly as a consequence of the greater holding power of the schools, factors which have not been clearly identified have led to a higher level of participation in mathematics learning at the pre-tertiary level, and a greater opportunity, in general, to maintain and develop skills in this field of study.

Problems in Sex Differences Research

Sex differences are a difficult area to study and it is necessary to draw attention to some of the problems of research that have no doubt led to a confounding of relationships and the contamination of results. First, there is the problem of developing tests in such a way as to eliminate items with

apparent sex bias. This poses problems in establishing what a test measures, and uncertainty in what a test measures causes problems in the comparison of results between different tests and different studies. Secondly, an experimental study that should involve the random allocation of students to a sex (treatment) group cannot be undertaken. Thirdly, there is the question of when are the observed differences of practical significance in contrast to statistical significance. In large studies, results are often found to be of statistical significance because of the large sample size yet the effect size is of little practical significance, but practical significance is difficult to define.

Outcomes of the Study

It is hoped that from this investigation evidence would be provided which would not only allow a greater understanding of the apparent sex bias in student performance on ASAT, but also, where necessary and desirable, alter the procedures employed in the construction of the test and its use. Changes could be made in the nature of the subscales, in the balance of the subscales, in the length of the test or in the types of items used. Alternatively, changes could be made to the scaling procedures employed, or in the methods currently used to assist students to prepare for ASAT. Such changes could be very worthwhile modifications to existing procedures and practices.

CHAPTER 2

A REVIEW OF PREVIOUS RESEARCH

Defining Bias

Since 1977, when the Australian Scholastic Aptitude Test (ASAT) was first used in the ACT as a moderating device, there have been differences in the performance of males and females on the test. This difference in mean group performance has been referred to as a "sex bias". There are various definitions of "bias" but all of the current definitions are to some degree similar. The following definition will be adopted for use in relation to ASAT.

> An item will be considered biased if examinees of equal aptitude, but from different groups, have an unequal probability of answering the item correctly.
>
> A test will be considered biased if examinees of equal aptitude, but from different groups, have an unequal probability of gaining any given test score.

If the test does not contain bias for or against any subgroup of the population, then students of the same aptitude will perform equally well. However, a difference in the mean scores of each sex does not necessarily indicate that the test contains a sex bias. It may be that the candidature is such that there are different distributions of aptitude within the single sex groups when students sit for ASAT. Unfortunately, no studies have been found, during a review of previous research, that have undertaken a comprehensive exploration of differences of performance between the sexes on aptitude tests or that have considered more than one aspect of the issue. Several facets of the problem of sex bias have been identified.

Differences in Abilities

A large number of research studies have been concerned with sex differences in ability. Substantial surveys of previous research were carried out successively by Anastasi (1958), Maccoby (1966) and Maccoby and Jacklin (1974). At the time of their review Maccoby and Jacklin concluded that the following differences had been reasonably well established. Girls have a greater verbal ability than boys. During the period up to early adolescence the sexes are very similar in verbal ability. However, after this the girls show a superiority in almost all aspects of verbal ability. Boys excel in spatial ability. The male advantage appears at about the same time that

the girls begin to excel in verbal ability.

Boys have a greater mathematical ability. Boys' mathematical skills develop more quickly than the girls' mathematical skills during the early adolescent period.

Explanations of the origins of sex differences in abilities have followed two streams. The first is concerned with physiological differences. These include brain lateralization differences, hormonal differences, genetic differences and maturational differences (Maccoby and Jacklin, 1974). Secondly, there are socio-cultural differences. These include the cultural values and societal norms by which a child is reared. They involve sex role stereotyping and differential treatment of the sexes.

Glickman (1976), after considering a variety of possible physiological factors such as those mentioned above, suggested that it was the socialization differences that were the more significant. Evidence from Keeves (1973) suggested that given the extent of variation between countries and school systems, social and cultural factors played a significant part in the development of sex differences, particularly related to achievement in mathematics.

Evidence from more recent research studies indicated that there is some uncertainty about the significance and size of sex differences. As the methodology of studies has changed so that the amount of training in particular areas is controlled (for example, Fennema and Sherman, 1977; Wise et al, 1979) sex differences in mathematical and verbal ability have no longer been consistently found. Some of the differences that had been observed would now appear to be explicable in terms of differential course taking.

Since the mid-1970s several studies have reported few or no differences on spatial tasks (Jacklin, 1979; Armstrong, 1981; Fennema, 1981; Senk and Usiskin, 1983). However, some sex differences in performance have shown up in recent studies. For example, males outperformed females on tests of problem solving (Armstrong, 1981), and the SAT-M (Benbow and Stanley, 1980). Senk and Usiskin (1983) have concluded that when experience was controlled, regardless of the difficulty or complexity of the items, girls and boys performed equally well. Nevertheless, it is still evident that in a survey of sex differences in intellectual abilities there remains much uncertainty (Murphy, 1977).

Differences in Preparation

A scholastic aptitude test such as ASAT is designed to be independent of variations in school curricula, subjects studied by students, and different amounts of preparation. Scores on ASAT should, therefore, not be dependent on courses studied by students or the specific coaching they received.

The effect of coaching for the Scholastic Aptitude Test (SAT) in the United States has been explored in a number of studies. Dyer (1953) tested more than 200 boys, in each of two comparable independent schools, who took the SAT in September 1951 and again in March 1953. In one school the boys received ten hours of coaching and in the other none at all. He found a small advantage for the coached group. The gains on the SAT-M were statistically significant but of doubtful practical importance, while the gains in SAT-V scores were of no practical significance. French (1955) reported on a study that involved classes from three public high schools. The students in one school were not coached, the students in another school were coached in the verbal area, and the students in the third school received coaching in both verbal and mathematical areas. The gains reported were statistically significant, but they were small in relation to the standard errors of the tests.

French also found sex differences in response to coaching for the SAT. Girls improved more on the SAT-V than on the SAT-M. In addition, girls taking mathematics courses at school improved more than the boys taking mathematics, while the girls not taking mathematics improved less than the boys not taking mathematics.

In a study by Dear (1958) ten schools took part in coaching sessions while students in the schools who did not receive coaching formed one control group and students in another nine schools formed a second control group. The gains in SAT-M scores were statistically significant but probably not large enough to give an unfair advantage to coached students. Gains on the SAT-V were not significant.

Frankel (1960) attempted to evaluate coaching programs offered by private coaching organizations. Frankel reported no significant gains in either the SAT-M or SAT-V scores. A similar study by Whitla (1962) found that small positive gains were made on both the SAT-M and SAT-V but neither were statistically significant. Coffman and Parry (1967) found that a course in accelerated reading increased SAT-V scores slightly.

The early research into the effect of coaching indicated that there was little to be gained from extensive preparation programs. However, Pike (1979) challenged the relevance of the early research to the current situation. Pike's review of research advocated the continuing investigation of the effect of preparation even though the majority of studies in his review indicated little benefit from coaching.

> ... the changing make-up of the examinee population needs examination. It is entirely plausible, for example, that more advantaged students represented in most of the studies cited in the College Board booklet, *Effects of Coaching on Scholastic Aptitude Test Scores* (most of them conducted in the 1950's), were already well prepared to do their best on the SAT to a degree that cannot be assumed for an increasing proportion of the current candidate population .... (Pike, 1979:5)

The controversy over the effects of coaching has continued since this review. In a critical appraisal of the effects of practice and coaching for performance on the Scholastic Aptitude Test (SAT), Slack and Porter (1980) said that various studies had shown that practice could make significant differences in performance on sections of the SAT. They reviewed a number of studies that illustrated that practice could improve scores. Moreover, they claimed that the Educational Testing Service (ETS) had tried unsuccessfully to explain these findings away, and they believed that studies sponsored by ETS or the College Board contained conclusions that demonstrated coaching to be ineffective even though the published experimental evidence indicated the contrary. In response, ETS replied through Jackson (1980). He claimed that Slack and Porter had not substantiated their arguments and had misrepresented the findings of many studies.

Bangert-Drowns, Kulik and Kulik (1983) carried out a meta-analytic study of the effects of coaching on achievement test scores. In the course of their review they compared the effects of coaching on aptitude tests to the effects of coaching on achievement test scores. They made the following conclusion about the effects of coaching on aptitude tests.

> In rough summary, one might expect an effect of coaching of approximately .45 standard deviations for the more coachable aptitude tests, and an effect of approximately .3 standard deviations for an average test, and an effect of approximately .15 standard deviations for the least coachable tests. (Bangert-Drowns, Kulik and Kulik, 1983:582)

Bangert-Drowns, Kulik and Kulik also reported that the optimum amount of coaching time for aptitude tests was not clear. While Messick and Jungeblut (1981) found a positive correlation between coaching time and scores on the

SAT, Bangert-Drowns, Kulik and Kulik suggested that practice and coaching programs should be brief.

The only existing Australian evidence was in relation to the Commonwealth Secondary Scholarship Examination (CSSE). Wilson (1969) found that candidates improved their scores on the science, mathematics, and humanities papers if given the opportunity to sit for them again after a period of six to eight weeks. A report from the ACER (1974) found that a program of coaching and practice could not be expected to produce substantial gains on scholastic aptitude tests such as the CSSE.

Although the early findings pointed to the fact that only small gains, if any at all, could be expected from coaching, evidence to the contrary has arisen from more recent studies. In this investigation it was not possible to determine the effects of preparation on ASAT scores. However, it was considered useful to determine the amount of preparation that students were undertaking for ASAT in the 1983 testing program.

Differences in Course Type

It is generally accepted that there are a greater number of boys than girls in Australian schools who study mathematics and science at the upper high school level (See Moss, 1982). Girls tend to study subjects in the humanities and social science areas. The reasons for this probably lie in the sex role stereotyping and socialization pressures within our society as discussed above. This might have led to the observed sex differences in performance.

As stated above, evidence has been presented suggesting that differential performance between the sexes in mathematics might be in part due to more boys than girls studying mathematics (Fennema and Sherman, 1977; Wise et al, 1979). As a consequence of this, several studies have sought to test the hypothesis that sex linked differentials in course taking accounted for sex differences on aptitude tests.

Benbow and Stanley (1980) examined sex differences in the quantitative section of the scholastic aptitude test (SAT-M). Their sample consisted of an extremely able group of 10,000 students who had scored in the upper two to five per cent on another standardized achievement test. They found no evidence to support the differential course work hypothesis.

The findings of Benbow and Stanley have been challenged on several grounds (Luchins and Luchins, 1981; Stage and Karplus, 1981; Tomizuka and

Tobias, 1981). The strongest criticisms have been concerned with the quality of the sample used and the ability to make generalizations from it. Since it was a sample of volunteers Tomizuka and Tobias argued that social pressures would have prevented the more able girls from coming forward and thus they might not have been adequately represented in the sample.

Pallas and Alexander (1983) also criticized the Benbow and Stanley study. They pointed out that the experience and achievements of a self-selected sample of extremely able students would not generalize to all students of a similar age. They considered that Benbow and Stanley had not made clear their intention to focus on mathematical reasoning. They also made a major methodological criticism:

> ... the Benbow-Stanley analysis provides at best only an indirect test of the differential coursework hypothesis. It does not, as would be preferred, compare the performance of boys and girls who differ in the exposure to the sorts of courses that are presumed to promote high levels of quantitative performance. Rather, it finds differences in performance in the absence of differential school experiences. (Pallas and Alexander, 1983:167)

Pallas and Alexander carried out a study that carefully detailed students' experience in mathematics in order to determine what the effects of differential coursework were on SAT-M scores. Their sample consisted of 6119 students in 24 public high schools from 15 communities in the United States. The students were all enrolled in fifth grade in 1961 and were tested successively in 1961, 1963, 1965, 1967 and 1968. Academic transcripts were used to code the courses students had taken according to course content. A performance in quantitative courses variable was also constructed to provide an overall grade point average, and to control for initial aptitude differences the School and College Ability Test (SCAT) was used.

The study found that male and female quantitative performance was not noticeably different at the beginning of high school. By the end of high school there were quite significant differences. Up to 60 per cent of this difference could be explained by differential course taking. They also found that girls who did take quantitative courses out-performed the boys. This increased the girls' average and therefore disguised some of the course work effect. However, Pallas and Alexander made no attempt to explain the source of the remaining differences in SAT-M scores but pointed out that the difference might have originated from socialization processes.

Benbow and Stanley (1983) replied to the criticisms of their 1980 study by Pallas and Alexander (1983). In defence of their study Benbow and Stanley said that they had clearly focused on mathematical reasoning and their conclusions had been incorrectly interpreted by Pallas and Alexander.

> Our sole conclusion was that differential course taking does take account of the sex differences observed. It could not account for the differences because we showed that the boys and girls in our study had the same amount of formal training. That is as direct as one can get. (Benbow and Stanley, 1983: 469)

They then concluded with the statement:

> ... the differential course taking hypothesis does not seem to be a viable explanation for the differences among the intellectually gifted. Among average-ability students the situation may be different. (Benbow and Stanley, 1983:471)

This would appear to clarify much of the debate about the sample that Benbow and Stanley (1980) had used and the generalizability of their results.

Benbow and Stanley (1983) also criticised the methodology of the study by Pallas and Alexander (1983). They claimed that the SCAT-Q was not a good proxy for the SAT-M in lower grades. Alexander and Pallas (1983) then replied claiming that Benbow and Stanley had not given an adequate reason against the use of the SCAT-Q as an input control.

The results of the Benbow and Stanley (1980) and the Pallas and Alexander (1983) studies should not be seen as contradictory. Benbow and Stanley used a sample of students with equal formal mathematical backgrounds. They found that there were differences in performance between males and females on the SAT-M. They therefore found that sex differences in performance on the SAT-M still existed after taking into account differential coursework.

Pallas and Alexander (1983) used a sample that had a wide range of mathematical backgrounds. After controlling for ability and exposure to mathematical courses they found a difference still existed between male and female performance. However, it was found that 60 per cent of the original difference could be explained by differential coursework. Pallas and Alexander had produced evidence to support the results of Benbow and Stanley that differences existed between the sexes for students with similar backgrounds. By using a sample that consisted of subjects with a range of

mathematical backgrounds and abilities they were able to explore the amount of difference that could be explained by the differential coursework hypothesis. This approach was not available to Benbow and Stanley because of the nature of the sample they had used.

Differences in Apprehension and Approach to Aptitude Tests

Increasingly, there has been an awareness in society and among educators that sex role stereotyping and social pressures has placed substantial restrictions upon the development of human potential. It has been believed that the social climate did not encourage girls to perform well in traditionally male areas (Horner, 1972; Stein and Bailey, 1973). In the area of mathematics previous research has suggested that a relationship existed between attitudes and performance. Husén (1967), Ryan (1968), Anttonen (1969) and Rounds and Hendell (1980) have all reported a positive correlation between attitudes towards mathematics and performance in mathematics. The reported correlations were in the range 0.2 to 0.4. Neale (1969) has argued that the presence of a positive correlation could not be used to support the claim that more favourable and confident attitudes towards mathematics caused an increase in performance. Fennema and Sherman (1977) have argued that there was evidence for such a causal link. They believed that lower confidence in success contributed to lower achievement scores.

Milton (1957) found that there was a positive relationship between sex role identification and problem solving skill. Carey (1958) tested the hypothesis that some sex differences in problem solving performance could be attributed to differences in attitudes towards problem solving. She found that females had less confidence in problem solving and there was a positive correlation between attitudes and performance for both males and females. She then attempted to change the attitudes of the students in special tutorial sessions. After these sessions, the females' developed a more positive attitude and performance increased. The improvement, however, was not significant for the males.

Hilton and Berglund (1974) concluded that there was a relationship between students' attitudes and their performance but the exact nature of the relationship could not be investigated.

> The problem is that the interaction between interest and achievement is probably instantaneous. One word of positive feedback from a respected teacher, guidance counsellor or peer, and the student is immediately more interested in mathematics and, as a consequence, immediately more able in performing mathematically. (Hilton and Berglund, 1974:234)

Aiken (1972, 1974) developed two scales: enjoyment of mathematics and value of mathematics. He found significant relationships between performance on SAT-M and his scales. Fennema and Sherman (1977), after controlling for mathematics background, found important relationships between socio-cultural factors and sex related cognitive differences. They identified a range of attitudes that girls had to mathematics and mathematics performance, and they concluded that:

> The pattern of differences in mathematics achievement, spatial visualization and affective factor variables strongly suggests the influence of socio-cultural factors. (Fennema and Sherman, 1977:69)

A more complete review of attitudes towards mathematics can be found in Aiken (1976).

The research outlined above pointed to the importance of attitudes and the effects attitudes might have on performance in mathematics. It was therefore considered necessary to explore this issue in relation to ASAT, in order to ascertain whether attitudes towards ASAT had an effect on performance on the ASAT test.

Item Bias

The study of item bias is an area that has grown in sophistication over the last few years. Early studies of item bias were concerned with content analyses of the items. Binet and Simon (1916) noted that an early version of their intelligence test had what might be called a "cultural bias". This was based on the observation that there were group differences between scores obtained by children of Parisian working class homes and those from higher class homes. Systematic investigation of the extent of cultural bias in standardized tests was not undertaken until about 1945. A classic study by Eells et al (1951) was the first to make explicit the concept of cultural bias.

Content analyses and differences in item difficulty have continued to be the method for examining item bias until very recently. Studies of the effect of content bias on the performance of males and females have been reported in only a few studies. Coffman (1961), Donlon (1973), Strassberg-Rosenberg and Donlon (1975) studied the SAT. Hicks and Donlon (1976) studied the Graduate Record Examination (GRE).

Coffman (1961) compared male and female performance on the March 1954 SAT-V. While total scores exhibited no differences, Coffman's research

indicated differences on individual items. On inspection of the items Coffman categorized them in the following way. Items involving words which described personal feeling or personality characteristics favoured females, while items using business or mechanical vocabulary favoured males.

Conlon (1973) corroborated Coffman's findings in a study of the May 1964 SAT-V and SAT-M. He found that items classified by the test constructors as Human Relationships, Humanities or Aesthetic-philosophical were easier for females, while items classed as World or Practical Affairs were easier for males. In his analysis of the mathematics section Donlon found that on average, males outperformed females by a very small margin. Only two of the 60 mathematics items demonstrated a difference in favour of females. Strassberg-Rosenberg and Donlon (1975) studied the April 1974 SAT and reinforced the findings of Coffman and Donlon.

Donlon et al (1977) studied a range of tests suitable for Grades 1-12. They found that females performed better on items which contained a female reference. The difference, however, was very small. They concluded:

> It is in one sense, reassuring to find the low level of statistical effects associated with content effects in this study. However, the findings must not be translated into a conclusion that attention to balanced references in tests is inconsequential. Tests are an important sampling of culturally significant behaviours. The condition of this sampling should reflect the values of the culture. The patterns of differentiation are in transition, and the older practices of more frequent references to males, of stereo-typical association and of sex and status linking should be abandoned. (Donlon et al, 1977:84)

More sophisticated techniques for the identification of item bias have been developed over the past few years. They have aimed at detecting items that behaved differently for subgroups of the population even when there were different distributions of ability within the subgroups. A variety of such techniques have been reported in the literature but few have been used to identify items that might have a sex bias.

Cordall and Coffman (1964) first reported the use of an analysis of variance technique during a study of the SAT, that tried to identify item bias in relation to racial differences. They found highly significant interactions between items and race group for both the verbal and mathematics subtests. Cleary and Hilton (1968) used a similar technique in a study of racial bias on the Preliminary Scholastic Aptitude Test (PSAT) and the SAT.

They found no significant interactions. The ANOVA procedure used could not identify specific items that might be biased, and as a result of this an alternative method called the transformed item difficulties (TID) was developed (Angoff, 1972; Angoff and Ford, 1973).

The TID method was used by Angoff and Ford (1973) to study item bias in the 1970 PSAT. They found a strong item by race interaction on both the verbal and mathematics sections of the test. They also found that the interaction decreased if groups were matched on a related ability. This indicated that the procedure was to some degree sensitive to differences in ability within each group. This was considered undesirable.

The next advancement in test item bias methodology was proposed by Scheuneman (1979). She used a chi-square technique that divided the groups into ability strata defined by score on the overall test. The proportions correct on an item in each strata were then compared across groups. This technique has been used by Scheuneman (1975, 1976, 1977, 1978). It has also received support from a number of comparative simulation studies (Ironson and Subovlak, 1979; Shepard et al. 1981). Scheuneman's technique has now been superseded by more complex chi-square procedures developed by Marascuilo and Slaughter (1981).

Perhaps the most sophisticated technique for the detection of item bias is the latent trait theory approach. It is the method that has the strongest theoretical support because the item characteristic curves that are used are sample invariant. Osterlind (1980) and Yeh and Corklin (1980) used the Rasch model (Rasch, 1960; Wright and Stone, 1979) in experimental studies that showed its feasibility as a technique for the detection of item bias. The three-parameter model (Birnbaum, 1968) has also received wide support but it is much more difficult to employ than the Rasch one-parameter model.

Differences in Retentivity

The effects of different retention rates on average group performance has been considered by Walker (1967). McIntosh (1959) suggested that the population of eligible students formed a pool of ability. The aim of any selection process was to take the cream of this pool. As a larger number was drawn from the pool, the average ability of those selected would decrease because of the finite nature of the pool.

Walker (1967) employed a mathematical model based on this idea in an attempt to explain differences in mathematics performance between countries. His moderate success indicated that applying the same principle to differences in ASAT performance could be useful if there were recognizable differences between the sexes in retention rates at a particular level of schooling.

Conclusion

The research surveyed has clearly indicated a range of factors that might be involved in the issue of sex differences in performance on ASAT. Various studies have looked at the effects of many factors but, in the main, each factor has been considered in isolation. Only Pallas and Alexander (1983) have been able to determine the proportion of the difference that could be accounted for by a single factor. By studying a number of factors in one study it might be possible to determine the contribution of each on the observed sex differences in ASAT performance.

CHAPTER 3

OUTLINE OF THE STUDY

Some Questions for Investigation

To a large extent the framework of the study was dictated by the evidence from previous research in the area. As discussed in the previous chapter, a review of research highlighted the separate importance of five key issues:

1 Retention
2 Attitudes
3 Preparation
4 Item Bias
5 Differential coursework

However, no previous research studies had analyzed all of these key issues concurrently in order to establish the relative importance of each in contributing to the observed differences in test scores. The present study, therefore, attempted to undertake this task by examining the above mentioned five issues simultaneously in order to establish the relative importance of each.

The study proposed the following questions:

1 Were there some properties of the candidatures that resulted in males appearing to perform better than females?

In the Australian Capital Territory a larger proportion of girls than boys remained at school. If these extra girls were less able then they would decrease the average performance of girls in general.

2 Were there differences between the sexes and the States in attitudes such as, expectancy of success and fear of success in responding to ASAT, which might have been related to sex differences in performance?

3 Were there differences between the sexes and States in the amount and type of preparation undertaken for ASAT?

4 Did the test itself have any properties that showed it to be biased against females?

This included an examination of the structure of the test and the content of the items.

5 Did students who studied particular subjects have an advantage over students who did not study those subjects?

Table 3.1 Mean Raw Scores on ASAT and its Subtests

			Total	Quantitative	Verbal	Humanities	Social Science	Science	Mathematics
1979	Australian Capital Territory	Males	61.8	31.4	28.8	17.9	11.9	20.0	12.0
	Queensland	Males	59.1	31.1	26.6	16.6	10.7	19.4	12.4
	Western Australia	Males	57.8	29.2	27.1	16.8	11.0	18.7	11.3
	Australian Capital Territory	Females	57.9	27.4	29.2	18.3	11.3	17.9	10.5
	Queensland	Females	55.0	26.7	27.0	17.1	10.3	17.1	10.6
	Western Australia	Females	52.7	24.1	27.2	17.1	10.3	16.1	9.1
1980	Australian Capital Territory	Males	72.2	40.5	29.8	20.2	13.0	23.5	15.5
	Queensland	Males	69.0	39.6	27.3	18.7	12.0	22.8	15.5
	Western Australia	Males	66.8	37.7	27.1	18.6	11.9	22.1	14.3
	Australian Capital Territory	Females	69.0	37.4	29.8	20.5	12.4	22.1	14.0
	Queensland	Females	66.1	36.4	27.8	19.3	11.5	21.3	13.9
	Western Australia	Females	63.9	34.1	27.8	19.5	11.4	20.4	12.6
1981	Australian Capital Territory	Males	63.0	32.6	30.3	17.3	12.8	18.7	14.2
	Queensland	Males	60.0	31.9	28.1	16.1	11.8	18.0	14.1
	Western Australia	Males	59.0	31.0	28.1	16.1	11.7	18.1	13.2
	Australian Capital Territory	Females	57.0	27.0	29.9	17.5	12.2	15.3	11.9
	Queensland	Females	55.4	27.0	28.4	16.7	11.5	15.2	12.0
	Western Australia	Females	53.8	25.6	28.2	16.7	11.2	15.0	10.9
1982	Australian Capital Territory	Males	62.9	34.5	25.6	16.9	12.6	20.3	13.1
	Queensland	Males	59.4	33.0	23.7	15.7	11.7	19.3	12.7
	Western Australia	Males	58.6	32.4	23.5	15.6	11.5	19.5	12.0
	Australian Capital Territory	Females	56.6	29.3	24.5	16.4	11.9	17.2	11.0
	Queensland	Females	53.4	27.7	23.0	15.4	11.2	16.1	10.7
	Western Australia	Females	53.6	27.4	23.6	15.8	11.2	16.6	10.0
1983	Australian Capital Territory	Males	63.2	36.2	26.7	17.0	11.8	21.2	13.2
	Queensland	Males	59.4	34.5	24.7	15.5	11.0	20.3	12.7
	Western Australia	Males	59.1	30.1	25.5	15.8	11.2	19.8	12.2
	Australian Capital Territory	Females	59.4	32.1	26.9	17.5	11.1	19.1	11.7
	Queensland	Females	55.7	30.4	25.0	16.1	10.4	18.0	11.3
	Western Australia	Females	55.9	30.1	25.5	16.7	10.7	17.8	10.9

Although performance on aptitude tests such as ASAT is supposed to be independent of the courses a student undertakes, it is clear that a student with a strong mathematical background may have less difficulty with an item based on mathematical principles. Similarly students undertaking humanities subjects will be more familiar with techniques that will help them answer humanities items.

Examining the Data

The study began with an extensive analysis of data files, containing raw ASAT data in order to establish the nature of the problem. These initial exploratory analyses asked the following questions:

1 Were there differences in the performance of males and female students on ASAT in all three examining regions?

2 Were there differences in performance of both male and female students on ASAT between the three examining regions?

3 Were there differences in the performance of male and female students on the ASAT subtests; i.e. Humanities, Social Science, Science, Mathematics, Quantitative and Verbal?

Table 3.1 reports the results of some of the preliminary analyses. Complete data on sex differences in performance of each subtest of ASAT were only available for the years 1979 to 1982. These data show that sex differences exist in all three examining regions. They also show that the increase that occurred in the ACT in 1981 and 1982 also occurred in the other two examining regions. In 1979 and 1980 the males outscored the females in the quantitative subtest and females outscored the males on the verbal subtest. In 1981 and 1982 males not only outscored females on the quantitative scale but they also tended to outscore females on the verbal subscore. Over each of the years, and in each authority, males tended to outscore females on the Science, Mathematics and Social Science subtests while females outscored males on the Humanities subtests.

In the above discussion of the differences between the States and between the female and male subgroups within each State, no attempt was made to examine the data recorded for statistical significance. In part this was because some of the samples were extremely large and even small differences in mean score would have been statistically significant. However, it is also important to recognize that in each of the State systems different proportions of each sex retained to the Year 12 level and sat for the Australian Scholastic Aptitude Test. The students in each State and of each sex group might be expected to differ in their average level of

scholastic aptitude as a direct consequence of the differences in retention rates for each group. It was also possible that, because of the structure of the Canberra society and the establishment of senior colleges in the Australian Capital Territory, the male and female student groups in the ACT might show characteristics that were not in common with those in Queensland and Western Australia. Consequently, it was considered important to replicate the examination of evidence obtained in the ACT by parallel studies in Queensland and Western Australia. If the findings were replicated in the three systems then a higher degree of generality might well be claimed for such findings.

After the nature of the problem had been established by these preliminary analyses, it was necessary to explore the problem within the framework of the five issues mentioned above. Two instruments were developed in order to do this, an interview schedule and a questionnaire. These are presented in Appendices 5 and 5 respectively.

The questionnaire contained items that obtained information on students' background, the courses they studied, the amount of preparation they did for ASAT and their attitudes towards ASAT. The interview schedule focused on student preparation, item type and style, differential coursework and the validity of ASAT.

The Key Issues

The five issues with which the study was concerned were restated in terms that could be experimentally explored.

Issue 1 Did the greater retention rate of females affect the mean of their ASAT scores?

This involved the development of a mathematical model that was used to predict performance differences when retention rates differed.

Issue 2 Were there sex or State differences in attitudes towards ASAT?

Information about expectancy of success, fear of success and apprehension in approach to ASAT were investigated by the questionnaire and the interviews.

Issue 3 Were there differences in the amount and type of preparation received for ASAT between the sexes, the three examining regions and school types within the three regions?

This issue was explored by means of the questionnaire and interviews which addressed the issues of the amount of home and school

preparation undertaken and whether students considered that this affected performance on ASAT.

Issue 4 Did the test itself have any properties that showed it to be biased?

Several statistical techniques for detecting item bias were employed using data files held at the ACER. It was possible to identify specific items or units within ASAT that exhibited differences in response patterns between the sexes. It was therefore possible to identify the characteristics of items that exhibited consistent differences in responses.

Issue 5 Did students who studied particular subjects have an advantage over students who did not study those subjects?

The questionnaire contained information about the amount of mathematics students studied, and the different subjects they studied. From this, it was possible to explore the relationship between ASAT and subject choice. This issue was also explored by the interviews which included questions on whether students perceived that particular subjects studied provided an advantage on ASAT.

Aims of the Study

The aim of the study was to determine if bias did indeed exist in ASAT within the framework of the five issues discussed above and if so, to attempt to explain its source and the relative importance of each issue in relation to that bias.

It was also expected that recommendations for policy and practice and the identification of areas of further research would arise out of the five issues explored by this study.

CHAPTER 4

RETENTION AND SCORE DIFFERENCES

Since the time when ASAT was first used in the Australian Capital Territory, sex differences in performance have been observed. During the period 1977 to 1982 there was an increase in the magnitude of these differences in performance. In 1983, the difference decreased but it was still considered substantial. During this same period the proportions of the age cohorts of both male and female students, remaining at the Year 12 level and sitting ASAT, have increased. Moreover, these increases have been greater for females than males.

It would seem possible that we might be able to explain the changes in the difference between the sexes in observed scores in terms of the changes in the differences of the relative proportions of males and females remaining at the Year 12 level and taking ASAT.

During the period 1977 to 1983 the total number of students sitting the ASAT examination has tended to increase while the absolute difference between the number of males and the number of females has remained relatively constant. However, there has been greater variation in the difference in proportions of the age cohorts remaining to the Year 12 level. These differences have increased over the seven year period. In 1977, 40.2 per cent of the male age cohort and 49.4 per cent of the female age cohort remained to Year 12 and sat for ASAT to obtain a Tertiary Entrance Score from the ACT Schools Accrediting Agency. In 1983, these proportions had risen to 44.1 and 54.3 per cent respectively.

Table 4.1 records the trends in retention rates for male and females over the years 1977 to 1983. The method of establishing the cohort size and retention rate is given in Appendix 4.1. The students who are considered to be retained are those students who remained at school until Year 12 and had a Tertiary Entrance Score calculated for them. They are referred to as the TE group. The observed differences in male and female mean ASAT scores for the TE group are presented in Table 4.2. It should be noted that the males consistently outscored the females over the seven year period. As these differences in retention rates have increased there remains the possibility that there has also been an increase in the number of less able females staying on at school. This would tend to reduce the mean score for female students.

Table 4.1 Proportion of Age Cohort Requiring Tertiary Entrance Score at Year 12

Year	Males			Females			Difference between Proportions
	Cohort	No at Yr 12	Propn	Cohort	No at Yr 12	Propn	
1977	1964	789	0.402	1837	907	0.494	0.092
1978	2066	853	0.413	1939	926	0.478	0.065
1979	2118	917	0.433	2003	1060	0.529	0.096
1980	2105	861	0.409	2017	986	0.489	0.080
1981	1990	843	0.424	2025	1082	0.534	0.110
1982	1981	844	0.426	1953	1069	0.547	0.121
1983	2065	910	0.441	1953	1061	0.543	0.102

The proposition to be examined in this chapter is that the observed differences between the mean ASAT scores of male and female students are associated with the differences in the proportions of the age cohorts of males and females who remain at school at Year 12 and take ASAT.

Alternative Models

To approach the problem of determining the effects of differences in retention rates it is necessary to make assumptions about the nature of the ASAT candidature. Depending upon the assumptions that are made, varying results may be obtained. These results must be explained in the light of the assumptions that have been made and the method of analysis that has been used. It is therefore necessary to present a range of methods that make slightly different assumptions.

Table 4.2 Student Mean Scores on ASAT

Year	Males	Females	Difference
1977	67.49	64.91	2.58
1978	68.84	65.83	3.01
1979	68.84	64.79	4.05
1980	68.98	65.79	3.19
1981	69.63	64.45	5.18
1982	70.27	63.48	6.79
1983	68.25	65.02	3.23

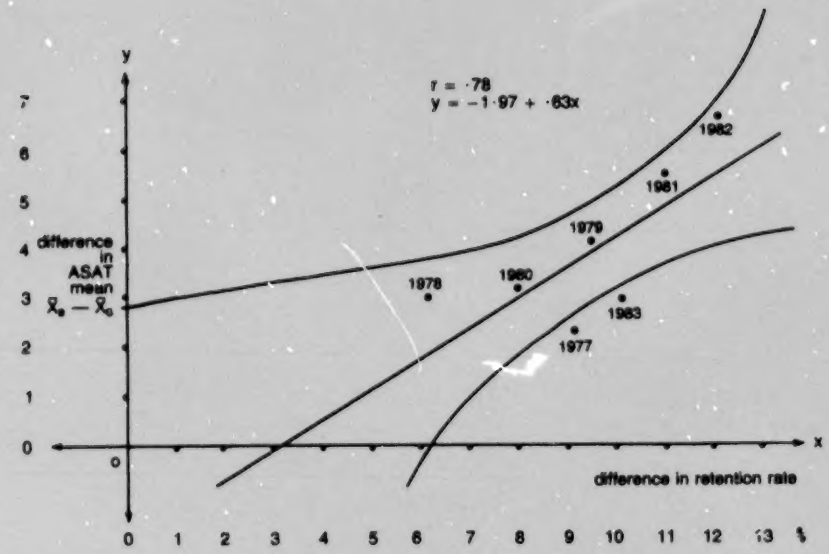

Figure 4.1 Plot of Retention Rate Differences and ASAT Score Differences

Method 1(a)

The most simple approach to the problem is to study the relationship between the difference in retention rates and the difference in ASAT performance using correlation techniques. The difference in retention rates and the difference in observed scores for each year were correlated using a Pearson product moment correlation. The correlation coefficient obtained was $r = 0.78$ ($p<0.05$). This indicated a strong relationship between these two variables considering the limited number of observations.

The regression of difference in score on difference in retention rates was $y = -1.97 + 0.63x$ (when both variables were expressed as percentages). The points are plotted and the regression line is drawn in Figure 4.1. This method predicted that if the retention rates were identical for males and females the expected difference would be -1.97 score units. That is, the females would outscore the male students. It was not possible to explain or interpret the slope of the regression line, 0.63, because of the totally different nature of the two variables in the regression relationship.

Table 4.3 Proportion of Male Age Cohort Retained with CGS Included

Year	Male Cohort Size	Males Retained	Proportion Males	Difference between Proportions (F-M)
1977	1964	909	0.462	0.031
1978	2066	973	0.491	0.007
1979	2118	1039	0.490	0.037
1980	2105	981	0.466	0.023
1981	1990	963	0.484	0.050
1982	1981	964	0.489	0.060
1983	2065	1030	0.499	0.044

It was also found that the origin was contained within the 95 per cent confidence bands about the regression line. This indicated the possibility that retention effects accounted for all the difference in observed scores. The -1.97 score units predicted that when retention rates were equal ($x=0$), the distributions of male and female ASAT scores differed. However, even if retention rates were identical the means of the scores might be expected to be different. Each year approximately 120 boys have attended Canberra Grammar School (CGS). These boys were included in the age cohort figures yet they were not retained in the ACT system. It might be reasonable to expect that these boys from a fee paying school would score above the overall male average on ASAT. The female students who could be considered as their counterparts were, however, included in the data. The inclusion of the CGS boys in those remaining to Year 12 would increase the retention rate for males and might be expected to increase the average male ASAT score.

Method 1(b)

Although it would not be possible to calculate the mean ASAT score for males with the CGS boys included, it would be feasible to include these boys in the retention rate data, on the assumption that they did not change the mean score of males on ASAT. Table 4.3 shows the revised male retention rates with the inclusion of the CGS boys. The retention rates for males were increased by about six per cent and the difference between male and female retention rates decreased by about six per cent.

After this modification the Pearson product moment correlation between score difference and retention rate was calculated to be $r = 0.78$, ($p<0.05$). The regression equation for the difference in score on the difference in

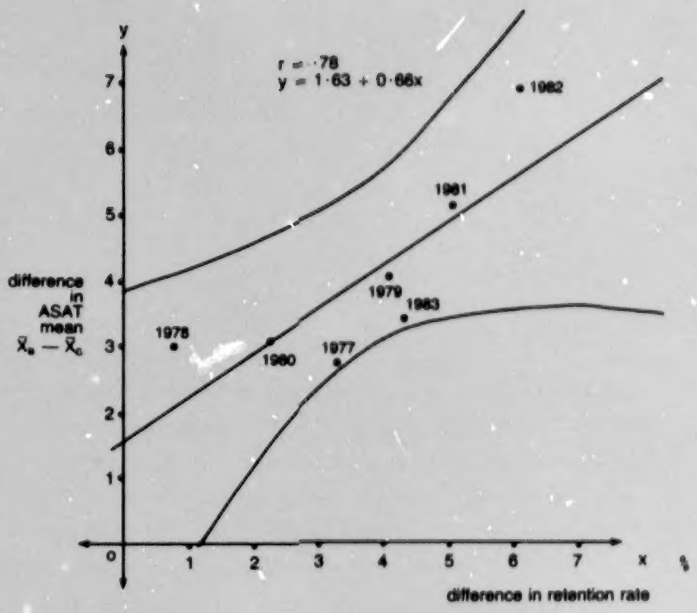

Figure 4.2 Plot of Retention Rate Differences and ASAT Score Differences with CGS Included

retention rate was y = 1.63 + 0.66x (when both variables were expressed as percentages). If the retention rate for males and females were equal (x = 0) the males would be expected to outscore the females by 1.63 score units. As with the previous method the slope of the regression line could not be interpreted. It was also observed that the origin was contained within the 95 per cent confidence bands about the regression line.

With the inclusion of the CGS boys it was found that the males would outscore the females if retention rates were identical. Retention might account for all of the difference because the origin was contained in the 95 per cent confidence bands about the regression line. The points recorded under the Method 1(b) are plotted and the regression line drawn in Figure 4.2. By including the CGS boys without altering the mean score of the males, the difference between male and female means was underestimated. Therefore, with CGS included the males might be outscoring the females by more than tha amount indicated by the regression equation.

Method 2(c)

The approaches used in Methods 1(a) and 1(b) did not indicate the amount of score difference that might be attributed to retention effects in each of the years 1977 to 1983. Another approach was available that provided an estimate of the amount of difference in male and female mean scores, that could be attributed to retention rate differences.

This method is best explained by an example. Suppose the male age cohort was 2000 and the female age cohort was 2200. In addition, suppose that 1000 males (50%) and 1200 females (54.5%) remained to Year 12. By taking the mean of the top 1100 female students (50% of age cohort) it would be possible to obtain an indication of the effects of the greater female retention rate. This method has been used previously by Husén (1967) and by Comber and Keeves (1973) in cross country comparisons.

In general, the modified mean for the females could be calculated

for the top $\frac{N_m}{N_{tm}} \cdot N_{tf}$ students

where N_m = the number of boys sitting ASAT

N_{tf} = the female age cohort

N_{tm} = the male age cohort.

This method can only be applied as described if the female retention rate is greater than the male retention rate.

The difference between the male mean ASAT score and the modified female mean ASAT score is an estimate of the difference that could not be explained by retention alone. The results in Table 4.4 show that when the modified mean was calculated, females outscored males in five of the years. These results were consistent with the first method which predicted that females would outscore males if retention rates were equal.

The adjusted mean female scores were, however, an overestimate. The method assumed that any extra girls who were retained were of lower ability than those that would normally be retained if the female and male retention rates were the same. It has been assumed that the greater retention rate was a result of lesser ability girls staying on. This of course, might not be the case. This would mean that the modified differences expressed in Table 4.4 were those still to be explained after retention differences had been accounted for, and would be greater in a positive sense than indicated; that is, they would be more in favour of boys.

Table 4.4 Comparison of the Difference between Male and Female Mean Scores and Difference between Male and Modified Female Mean Scores

Year	Male Mean	Female Mean	Modified Female Mean	Observed Difference	Modified Difference (M - F)
1977	67.49	64.91	70.25	2.58	-2.76
1978	68.84	65.83	69.34	3.01	-0.49
1979	68.84	64.79	69.89	4.05	-1.05
1980	68.98	65.79	70.27	3.19	-1.29
1981	69.63	64.45	69.46	5.18	0.17
1982	70.27	63.48	69.04	6.79	1.23
1983	68.25	65.02	69.86	3.23	-1.61

The consistency of the results from Methods 1(a) and 2(a) should be noted. They both support the notion that when retention differences had been accounted for, females could be expected to outscore males. However, if retention rates within the ACT system were to be equal a greater proportion of the male cohort than the female cohort would have remained to Year 12 because of the boys attending CGS. This method has indicated that if retention rates were equal, without taking account of CGS, more lower ability males than females might be retained. This could be due to the higher true retention rate (i.e. proportion of cohort undertaking a Year 12 course) for males, or the exclusion of CGS, which might underestimate the males' mean score. The inclusion of CGS could also increase the male retention rate and therefore decrease the modified female mean. This approach must be considered in an attempt to explain why females might be outscoring males.

Method 2(b)

The formula used Method 2(a) could be altered to:

$$\frac{(N_m + 120)}{N_{tm}} \cdot N_{tf}$$

where N_m = the number of boys sitting ASAT

N_{tf} = the female age cohort

N_{tm} = the male age cohort.

The inclusion of 120 males is an estimate of the number of boys attending CGS.

Table 4.5 Difference between Male and Modified Female Mean Score with the Inclusion of Canberra Grammar School

Year	Male Mean	Female Mean	Modified Female Mean	Observed Difference	Modified Difference
1977	67.49	64.91	66.57	2.58	0.92
1978	68.84	65.83	66.32	3.01	2.52
1979	68.84	64.79	67.28	4.05	1.56
1980	68.98	65.79	67.36	3.19	1.62
1981	69.63	64.45	66.93	5.18	2.30
1982	70.27	63.48	66.43	6.79	3.84
1983	68.25	65.02	67.31	3.23	0.94

With the inclusion of the CGS boys, the modified female mean was decreased. The differences between male and female means were therefore increased. It might, however, still be considered an underestimate because the mean of the males was not increased by the inclusion of CGS, as might be expected. The differences between male and modified female mean scores with the inclusion of CGS are presented in Table 4.5.

The data in Table 4.5 show that a substantial part of the difference in mean scores can be attributed to differences in retention. If the inclusion of CGS were to increase the male mean by only a small amount, this method indicates that in years 1979 to 1983 about 50 per cent of the observed differences in mean scores may have been due to retention differences. If the inclusion of CGS increased the male mean by a large amount, then there would be a substantial increase in the amount of score difference still to be explained.

As with Method 2(a) the calculation of the modified female mean assumed that any extra females were of lesser ability. This overestimated the modified female mean and in this case left a greater amount of the difference still to be explained. The magnitudes of differences still to be explained as indicated in Table 4.5 were clearly underestimated as they did not take into account these two considerations.

Discussion

These four approaches have illustrated that differences in retention rates might be playing a considerable part in the observed differences. Methods 1(a) and 2(a) indicated that if CGS were ignored then females would be

outscoring males, when retention rates were considered equal. From the confidence bands around the regression line in Figure 4.1 it was, however, possible to say that retention differences might account for all the observed score differences. Methods 1(a) and 2(a) showed that the distribution of ability among the students sitting ASAT was different for males and females. The difference in score in favour of the females might therefore be accounted for by the following explanations. First, when retention rates are considered equal, the retention rates are those within the ACT Schools Authority system with the males attending CGS not retained. Therefore if the retention rates for males and females were set to be equal then the true retention rate for males would be greater than that for females. This might explain why females could be expected to outscore males. Secondly, it is generally acknowledged that some more able and middle ability males leave school to obtain places in technical colleges or full-time employment while their female counterparts remain at school.

> It seems that some boys choose to leave school earlier than they would otherwise have done in order to obtain employment. This tendency has been reinforced by unemployment benefits which, for the young, increased sharply in 1973 and 1974. The contrary trend for girls is possibly related to the lack of full-time employment opportunities for them combined with, perhaps, a greater reluctance of parents to see their daughters on the dole. (Karmel, 1981: 14)

This possibility is ignored by Methods 1(a) and 2(a) because these methods make no allowance for different retention patterns. The expectation of females to outscore males only indicated that a greater proportion of middle and upper ability males were not being retained. Under these circumstances fixing retention rates to be equal therefore required the retention of an increased proportion of less able males. Methods 1(b) and 2(b) come a step closer to the real situation by allowing for the inclusion of the CGS boys. These methods indicated that if the male students attending CGS were included in the retention rate figures and if it were assumed that they did not change the mean ASAT score of the males, then the males would be outscoring females if retention rates were equal. If the CGS boys were to add to the male mean ASAT score then the amount by which males outscored females would be increased. Method 1(b) showed that all of the differences might be attributed to retention while Method 2(b) suggested something just below 50 per cent of the observed differences might be accounted for by retention.

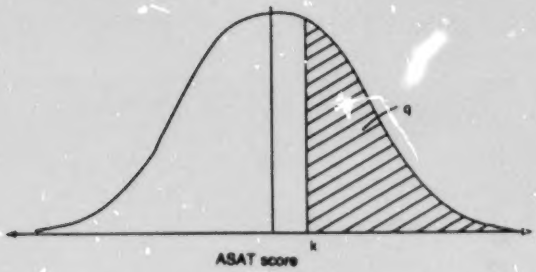

Figure 4.3. A Hypothetical Distribution of ASAT Scores for the Age Cohort, showing the proportion q above the cut-off point, k.

Model 3(a) A Model for the Effects of Selection

Following an approach first suggested by McIntosh (1959) and developed further by Walker (1967) it was possible to transform the retention rates into expected scores in a standard score form. This allowed the development of a model to account for the effects of selection. A set of assumptions were necessary. The three most basic assumptions were:

1 Students' scores on ASAT would have been normally distributed over the full age cohort of boys and the full age cohort of girls if they had remained at Year 12.

2 These two hypothetical normal distributions were identical.

3 Only the "best" students remained at Year 12.

The normal curve in Figure 4.3 represents the expected performance on ASAT if the full age cohort of males and females had remained at Year 12. This distribution is assumed to be identical for the male and female age cohort. The shaded region to the right of the cut-off point, k, represents those students who actually remained until Year 12. Since a greater proportion of the female age cohort remained the cut-off point would be further to the left for females than males. On the basis of this model we can calculate the expected mean scores for both the males and females for each of the years 1977 to 1983. Derivations of the formulae are given in Appendix 4.2.

Expected mean = y/q

Variance = $1 - y/q(y/q - k)$

where q = proportion of cohort who remain

y = the ordinate of the normal curve at the cut-off point k.

Table 4.6 Expected and Observed Differences in ASAT Scores

	Expected				Observed			
Year	Male	Female	Difference	Standard Difference	Male	Female	Difference	Standard Difference
1977	0.962	0.809	0.155	0.264	67.49	64.91	2.58	0.186
1978	0.943	0.833	0.110	0.184	68.84	65.8	3.01	0.214
1979	0.908	0.752	0.156	0.258	68.84	64.79	4.05	0.293
1980	0.950	0.816	0.134	0.239	68.98	65.79	3.19	0.229
1981	0.924	0.744	0.180	0.295	69.63	64.45	5.18	0.371
1982	0.919	0.724	0.195	0.322	70.27	63.48	6.79	0.486
1983	0.896	0.731	0.165	0.271	68.25	65.02	3.23	0.231

The differences between these scores can be calculated and compared to the observed differences in scores.

In each of the years 1977 to 1983 the students' ASAT scores were standardized to a fixed mean and standard deviation. However, since the TE group was only a subgroup of this, the standard deviation of the TE group's scores varied from year to year. An approximate standard deviation for the TE group was 14.

We can therefore transform the observed differences in ASAT means to approximate standard scores by dividing by 14. The expected scores can be transformed to standard scores by dividing the differences by the standard deviation of the male and female students combined. For the derivation of the standard deviation of the pooled population of males and females see Appendix 4.2. Table 4.6 records the calculated values of the standardized expected differences and the standardized observed differences in scores. The Pearson product moment correlation between the standardized observed and standardized expected differences was $r = 0.75$, $(p<0.05)$. The regression equation for the difference in observed score on difference in expected score was, $y = -0.187 + 1.82x$. This means that if the retention rates were identical we would expect the females to outscore the males by 0.187 standard score points. The points are plotted and regression line shown in Figure 4.4.

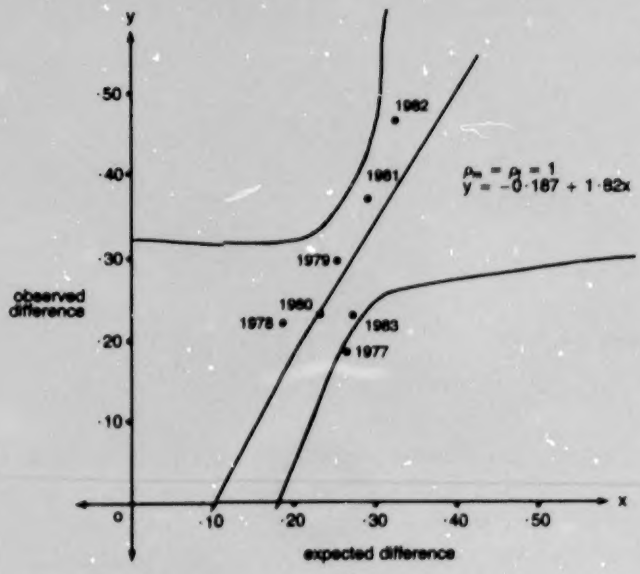

Figure 4.4 Plot of Expected Differences in ASAT Score Against Observed Differences in ASAT Score

As with Methods 1(a) and 1(b) the origin is contained within the 95 per cent confidence bands about the regression line. Retention might therefore be accounting for all of the observed score differences. The correlation of 0.75 is marginally lower than the 0.78 that was achieved with Method 1(a). This is a consequence of the assumptions required for the calculation of an expected score. While the gradient of the regression line in Method 1(a) had no meaning the regression line in this method indicated the observed score differences were about two times as large as expected. This is indicated by the gradient, b = 1.82, of the regression line.

Although this model has the advantage of calculating an expected score difference, it makes unrealistic assumptions about the distribution of ASAT scores. Assumption 3 is clearly an oversimplification of the actual process involved in selecting students who remained to Year 12 and sat for ASAT in order to obtain a TE score. In effect, Assumption 3 states that the 50 per cent of students sitting ASAT were the 50 per cent of students who would have achieved the highest scores if the whole age cohort had sat ASAT. If Assumption 3 were correct the distribution of ASAT scores for male and female students would be shaped like the shaded region in Figure 4.3.

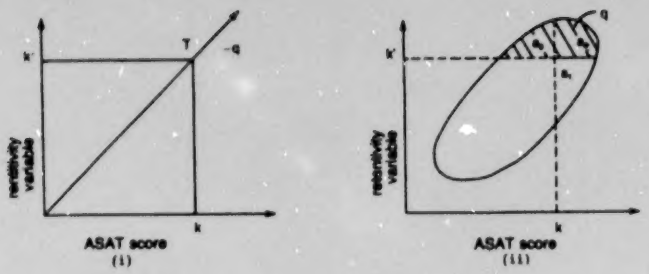

Figure 4.5 Correlation Surface Relating ASAT Score to the Retentivity Variable

Method 3(b)

This oversimplification could be overcome by allowing a correlation, ρ, between the variable operating to select students who remained until Year 12, which might be called the retentivity variable, and the ASAT score. This correlation could be different for males and females, but in both cases it might be expected to be large. The special case of $\rho = 1$ for both males and females would give the model in Method 3(a). Figure 4.5 records the correlation surfaces relating the retentivity variable to the ASAT score distribution.

In the original model the proportion of the age cohort remaining at Year 12 would lie above T on the straight line in Figure 4.4(i). When perfect efficiency of selection is not assumed Figure 4.4(ii) illustrates the relationships between ASAT score and the variable determining selection. The regions marked a_1 and a_2 make up the proportion, q, that remained to Year 12. Those students in the region a_1 would have achieved above the cut-off score, k, if they had remained, while those students in the region a_3 remained at school yet achieved below k. The regions a_3 and a_1 would contain equal numbers of students if the cut-off score, k, corresponded to the cut-off point of the retentivity variable, k'. As ρ increases these regions become smaller.

If we assume that the retentivity variable and ASAT score have a joint normal distribution, the expected mean score becomes $\rho y/q$, where ρ is the correlation between the retentivity variable and ASAT score and y and q are defined as in Method 3(a). The variance is $1 - \rho^2 y/_q(^y/_q - k)$. These results are derived in Appendix 4.2.

By varying ρ for males and females it was possible to allow for different retention patterns for the sexes. This would provide compensation for the CGS boys, a problem which until now we have not been able to deal with to our satisfaction. Since the true values of ρ are not known, a procedure was employed which allowed ρ to vary independently between 0.6 and 1.0 for the males and females. Table 4.7 shows the correlations between the expected difference in ASAT scores and the observed differences in ASAT scores when ρ varies between 0.6 and 1.0. Given that the CGS boys were not included in the retention rate it could be expected that $\rho_m < \rho_f$. This is borne out by the non-significant correlations in the bottom left of Table 4.7. For the values on the diagonal in each of the years 1977 to 1983, males were expected to outperform females. The differences predicted by the model were, however, somewhat less than the observed differences. For $\rho_m = \rho_f = 0.6$ the expected differences were about one quarter of the observed differences. For $\rho_m = \rho_f = 1.0$ the expected differences were about two thirds of the observed differences. For $\rho_f > \rho_m$ the results of the model are sometimes difficult to interpret. It is clear that in this region the model fits the data best in terms of significant correlations. However, for the values above the dotted line, marked in Table 4.7, the females were expected to outperform the males in at least one of the years 1977 to 1983. This would occur if ρ_m were allowed to be significantly less than ρ_f. In this situation the distribution of male ASAT scores would be spread further down, therefore making the males' mean less than the females' mean. This implies that the model might support the notion that the different retention rates were disguising a greater difference between male and female ability than has been currently observed. This is unlikely to be the case since ρ_m and ρ_f should be reasonably close together. Although the model fitted the data best in this region it is not significantly better than the region between the dotted line and the diagonal. (See Table 4.7). In this region both ρ_m and ρ_f are high, which is more likely to be the case, with $\rho_f > \rho_m$. Furthermore, the expected score differences favoured the males. The expected differences were, however, only between about two thirds and one half of the observed differences.

The results in Table 4.7 show that $\rho_m > \rho_f$ does not fit the data. For $\rho_f > \rho_m$ the data fit the model but provide a variety of possible situations. If $\rho_m = \rho_f$ were accepted, then it could be suggested that approximately one half to one quarter of the observed differences would be explained by retention. As we allowed ρ_f to become greater than ρ_m the data fitted the model more closely, however, the amount of difference explainable by

Table 4.7 Correlation Between Observed and Expected Differences in Scores for the Modified Model

ρ_m \ ρ_f	0.6	0.7	0.8	0.9	1.0
0.6	0.74	0.77	0.77	0.76	0.76
0.7	0.75	0.77	0.77	0.77	0.76
0.8	0.70	0.75	0.76	0.77	0.77
0.9	0.62	0.71	0.74	0.76	0.77
1.0	0.48	0.62	0.70	0.74	0.75

retention decreased. For large differences between ρ_m and ρ_f with $\rho_f > \rho_m$ the females could be expected to outscore the males.

The regression line of difference in expected score on difference in observed score for $\rho_m = 0.8$ and $\rho_f = 0.9$ is shown in Figure 4.6. For these rho values the observed differences are 2.21 times larger than expected and the intercept of 0.143 indicated that males would have outscored females when the model predicted no difference.

Summary and Conclusions

It is clear that retention effects must be considered when looking at differences in ASAT mean scores. Methods 1(a) and 2(a) showed that females might have outscored males after retention differences were accounted for. These two methods ignored the possibility of different retention patterns for the two sexes. They also ignored the boys who attended Canberra Grammar School. Methods 1(b) and 2(b) attempted to allow for Canberra Grammar School but could not be completely successful because it was impossible to determine the effect of including Canberra Grammar School on the mean of the males' ASAT scores. These two methods did, however, indicate that somewhere between half and all of the observed score differences might be explained by retention differences.

Method 3(a) transformed the retention rate differences into expected score differences thus enabling a greater understanding of the relationship between retention differences and score differences. This method did not allow for different retention patterns. It was found that the observed score differences were about one and a half times as large as expected and

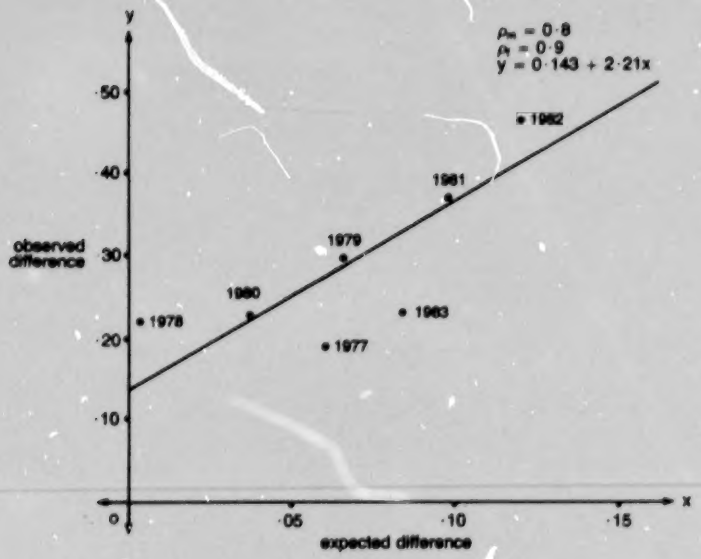

Figure 4.6 Plot of the Expected Difference Against the Observed Difference $\rho_m = 0.8$, $\rho_f = 0.9$

females would have been expected to outscore males if retention rates were equal. Alternatively, it could be said that retentivity effects might be expected to account for about two thirds of the observed score differences.

Method 3(b) was introduced because it allowed for different retention patterns. This included an allowance for Canberra Grammar School. The model most accurately fitted the data for $\rho_f \rightarrow \rho_m$. If both ρ_m and ρ_f were taken to be large then the model predicted that male students would have outscored females by 0.143 standard score units when retention effects were taken into account. This residual might be explained in terms of the following three factors: (1) curricular differences through different subject choices for males and females, (2) test and item bias, and (3) differences in attitudes and expectations between the sexes. The gradient of the regression line in Figure 4.6 is more difficult to interpret. It appears to indicate a relationship between the three factors mentioned above and retentivity. The model has shown that retention might be able to account for about half of the observed differences in ASAT scores.

It should also be noted that the models that we have proposed for retentivity effects accounted for about half of the variance $(0.75)^2$ of the observed scores across the seven years from 1977 to 1983.

CHAPTER 5

INTERVIEWS WITH STUDENTS

Interviews with 56 students who completed ASAT were carried out in Queensland and the ACT. It was not possible to interview Western Australian students as it was felt that it would disturb their examination and study timetables. The sample consisted of 56 students (27 male, 29 females). Thirty of these were from Queensland (15 male, 15 female) and 26 were from the ACT (12 male, 14 female). In the ACT two Year 12 students from each of the colleges offering Year 12 courses were selected. One college, Narrabundah, was closed at the time and could not be visited. In Queensland, two Year 12 students from each of 15 schools were interviewed. The Queensland schools were a subsample of the questionnaire sample selected by the ACER.

In each of the co-educational colleges one male and one female student were interviewed. All interviews with female students were carried out by a female interviewer. Different female interviewers were employed for this purpose in the ACT and Queensland. Most interviews with male students were carried out by a male interviewer, who was the chief investigator for the study. The interviews were recorded and only very brief written notes were taken during the interviews. An interview schedule was designed to obtain information relating to the following six areas:

A Do students feel that preparation and coaching make a difference to their mental state and/or performance on ASAT?

B Do students consider that those students with experience in mathematics and science subjects have an advantage on ASAT?

C Do girls approach ASAT with greater apprehension?

D Do students consider that boys have an advantage on ASAT because
(a) of a greater willingness to engage in risk taking behaviour, or
(b) the item format favours the boys?

E Do differences exist between the sexes in their perceptions of whether or not ASAT is a valid indicator of scholastic aptitude?

F Are girls more likely to be disadvantaged for physical or health reasons?

After the ACT interviews some minor alterations were made to the schedule. They included changes in item order and the elaboration of some items that required further explanation. The revised interview schedule is presented as Appendix 5.1. After carrying out the interviews it was considered that the major issues of concern did not fall neatly into the original categories as proposed. To make the discussion below more meaningful the following topic areas have been identified and used.

A Students' understanding of the use of ASAT
B Differences in preparation and coaching between schools.
C Apprehension levels of students.
D The effect of course type on performance.
E Strategies employed in working through the test.
F Validity of ASAT as perceived by students.
G The possibility of sex bias.

In addressing the issues under these headings it was believed that the students' views and impressions would be more accurately represented. However, it should be noted that all of the points in the original six categories have been covered within the six revised topic areas.

Australian Capital Territory

Debate regarding the quality of the educational system in the ACT was widespread within the territory. There had been considerable community and press involvement in discussing the relative merits and weaknesses of the system. Furthermore, the ACT Schools Accrediting Agency had made several public statements about the method in which ASAT was used to moderate school-based assessments. For students in the ACT, ASAT had a dual purpose. In addition to being used for the moderation of school assessments, a student's ASAT score also appeared on the Certificate of Supplementary Information for Tertiary Entrance. The interviews of ACT students were carried out in a period of four to six days after they sat ASAT.

A Students' understanding of the use of ASAT

Of the 26 students interviewed in the ACT only eight said that they were familiar with the details of the method in which ASAT was used to moderate their school assessments. The most typical response from students when asked about what they thought of the use of ASAT was summed up by the following comment from a student of a co-educational Government college:

> "I am not really knowledgeable on the workings of the ASAT test and the things behind it."

This later extended to:

> "I am not really sure what ASAT is used for. It is used to standardize your marks, isn't it?"

While most of the students were aware that ASAT was used in the standardization procedures, nine of the students had no idea of how it was used:

> "It is used to see if you can get into University."

This statement illustrated the level of understanding of a male student from a mixed Government college. As a consequence of his ignorance he was very worried that if he did not perform well on ASAT he would fail Year 12. Another boy from a boys' Catholic school also believed that:

> "... if you go to an employer or a Uni or CCAE (Canberra College of Advanced Education) they have to have some idea how a student has fared and the ability to handle work and stuff like that."

While most of the students confessed that they did not understand the details of the use of ASAT, 18 of the 28 students said that they were aware that the average mark of the group was the most important thing. According to the students interviewed, teachers had emphasized that the group mean and standard deviation were the most important factors involved in the use of ASAT for scaling purposes. Although this was correct and should have been explained to students in order to avoid misconceptions similar to those expressed above, it should also have been made clear that the way to increase the mean would be for every student to try his or her hardest. Judging from students' responses to interview questions, this was not necessarily the case. Students were of the opinion that the effect ASAT had on their marks was out of their control and consequently they did not always try as hard as they should have. There would appear to be a definite need for the students to have the workings of ASAT explained to them more fully so that they could approach the test with attitudes that would make it best serve them.

B Difference between Schools in Preparation and Coaching

All the students interviewed from non-Government schools indicated that the preparation that they had received was sufficient, in terms of both practice at items and information about the use of ASAT. They felt it had been particularly helpful in allowing them to understand the structure of the test.

Girls from two of the single sex schools indicated that they had been given more than one trial under examination conditions. The source of material for the extra trials was not made clear by the students. These girls had also been given special sessions involving practice at problem solving.

> "We had about two hours of problem solving on a Friday afternoon for about three weeks before ASAT. This made me prepared for what to expect, how quickly I had to work."

It was also found that the non-Government schools had provided their students with more adequate explanations about the details of the ASAT moderation procedure. The possible effects that it could have on their results had also been made clear to them.

> "... If you get a good mark then the average of the class goes up and you can get an extra 5 or 6 marks that pass you and allow you to get into Uni. This didn't build up nerves but it put a little bit of pressure on that I had to go well on it. As well as I can. Instead of putting 100 per cent effort I had to put 110 per cent effort."

The important fact that a student's ranking could not be changed by ASAT was only expressed and understood by students from one Catholic boys college.

> "As a whole I was not particularly nervous about it. The fact that your individual ranking won't be changed by it probably influenced the fact that nobody really worried about it."

The same student then said that it had been very clearly stressed to the students that their individual ranking within the class would not be changed by ASAT.

It was found that the preparation materials were generally available in the Government schools but students did not necessarily use them. It is convenient to divide the students at Government colleges into three categories with respect to the amount of preparation they had had:

1 those who did little preparation but still felt confident;

2 those who had prepared and felt confident; and

3 those who had prepared but did not feel confident.

There appeared to be approximately equal numbers of interviewees in these categories.

Category 1 students felt that preparation would be useless for ASAT because of its design. They did not feel that familiarization with the format would be particularly useful.

Category 2 students had all done the trial ASAT and studied the Student Information Bulletin. They had felt ready to tackle ASAT.

Category 3 students had also used all the preparation materials available but still had found ASAT particularly difficult.

A typical Category 1 student did not attempt to seek out the facilities provided by the school for preparation. The Student Information Bulletins were often distributed from the school office, and students had to collect them if they were interested. However, many did not. Some optional preparation sessions were provided.

> "Well, personally I didn't do any, but over the holidays two four- or five-hour sessions on methods for answering multiple choice questions and how to save time were given."

A similar view was expressed by a student at another school. He felt confident about ASAT and appeared to be quite able:

> "Well they did give us a booklet but I [illegible] I was assured by most people that it didn't really help you much."

Category 2 students completed the ASAT trial and worked through the Student Information Bulletin. They felt that this preparation was very good. They believed it helped them to develop strategies such as time allocation and attempting easier items first.

> "I decided after the trial and the information bulletin how I would do my ASAT. I am not very good at English and I get uptight about the English. So I knew when I got to do the real ASAT I decided I would leave all the English to last, do all the others and come back to the English."

Category 3 students had sought information like Category 2 students yet they were not satisfied with what was provided.

> "I went through a few of the questions and thought gee these are elementary, no worries here. Then I actually got to ASAT and got a big shock."

These students did not find the advice from teachers useful and said they were not provided with sufficient information.

> 'On the day of the test one girl actually put up her hand and said 'if you answer a question wrongly do you get marks taken off?' That's how much she knew about it."

When asked if she felt that adequate preparation was available one girl said:

> 'No-one knows anything about ASAT, I regret to inform. People just go blindly in during Year 12."

She felt that students and teachers alike did not know the use of ASAT and that more information should be available.

The non-Government schools provided a lot of preparation for ASAT concerning both its structure and its possible effects on school assessments. Government schools provided information on the structure of the test (if students chose to use it) but provided little information about the effects that scaling procedures might have on their assessments.

C Apprehension Levels of Students

In general, students appeared to have approached ASAT with a similar degree of apprehension. They were resigned to the fact that they had to do it. Only one student, a girl, was clearly distressed by the experience. The female students interviewed did not express a tendency to be more nervous and apprehensive than their male counterparts. Of the 14 girls interviewed, four said they were nervous. Two boys out of the 12 said that they were particularly nervous.

An interesting comment came from one of the girls:

> "I personally found it a good exam for myself, but in the group of people I know it scared a lot of people. Mainly the ones who are doing the vocational subjects; textiles, cooking, etc."

This comment suggested that course type might be the key to decreased confidence. Of the 10 mathematics and science students interviewed only one reported having had any apprehension about doing ASAT. Of the 16 others, six said that they were worried and apprehensive about doing well. It would appear that it might be the humanities students who faced ASAT with greatest apprehension because, it would seem, they had had less experience with multiple choice items and formal examinations.

D The Effect of Course Type on ASAT Performance

It was clear that the girls believed that studying mathematics and science gave a clear advantage when it came to doing well on ASAT. Their opinions did not seem to be affected by the type of course that they themselves were taking. The boys, however, were evenly divided between mathematics and science study as an advantage and study in English being considered an advantage. The boys who did mathematics and science courses felt that studying higher levels of English was advantageous.

> "To be good at English was particularly useful because the language of the exam is particularly sophisticated and many students had trouble understanding it."

> "In the English type questions you can interpret things in many ways but if you have got a lot of English you can interpret in the right way."

Table 5.1 Use of Complex Strategies by ACT Students in Answering ASAT.

	Yes	No	Total
Maths/science	4	6	10
Others	5	11	16
Male	5	7	12
Female	4	10	14
Coeducational	6	10	16
Single sex	3	7	10
Total	9	17	26

The boys doing the humanities subjects believed that studying mathematics and science would definitely be an advantage. Along with the girls who said that mathematics and science was an advantage, these boys said it was the method of training rather than the mathematics and science knowledge that gave the advantage.

> "Maths and science prepare you more logically. They are not creative."
>
> "Mathematics and science subjects give you more opportunity to do problem solving. Not so much your math skills or your chemistry skills but you develop skills in problem solving which you wouldn't get if you studied textiles."

Other students said that they considered they needed a refined and analytical type of reasoning power to answer the ASAT questions and they believed it was the type of reasoning that was developed in mathematics and science subjects that was relevant.

E Strategies Employed in Working through the Test

Of the 26 students interviewed only six used no particular strategy at all in answering the items, apart from working straight through the paper doing each question as they got to it. If they did not finish all the items in time, they guessed those that they had not yet reached. The other 19 students allocated time to each item and guessed if they exceeded the time allocated. More complex strategies such as doing a particular item type first were used by very few students. In Table 5.1 information is recorded on the use of complex strategies by ACT students in answering ASAT. Table 5.1 shows that nine students used what may be considered to be complex strategies. Among the students interviewed the use of complex strategies was not more widespread amongst any of the subgroups shown in Table 5.1.

There were no differences between boys and girls or between school type as to whether or not students allocated time to items. It was, however, observed that all 10 mathematics and science students allocated time to each item, whereas, six of the other 16 students did not do this.

F Validity of ASAT as Perceived by Students

All of the students interviewed believed that a system of school-based assessment was better than external examinations. All students believed that moderation was necessary to standardize assessments across schools and subjects. However, some students believed that there were better methods than ASAT available for such moderation. One student suggested separate examinations for each subject since:

> "It (ASAT) cannot be viewed as an accurate assessment of a student because of the different subjects that they study."

It was also suggested that no moderating instrument was necessary. There should be more communication between teachers in the different colleges to ensure they were setting the same standards. This type of communication could include things such as the exchange of pieces of student work to compare assessments in a rudimentary form of consensus moderation.

The majority of students said that they could not think of anything better than ASAT, yet they saw it as having limitations. When discussing ASAT's ability to measure a student's personal standard of education the most common complaint was that it could not test creativity.

> "It doesn't test my education. It ignores creativity in the first place. I did an enormous painting this term and I would like to see ASAT mark my painting. I don't think it could."

> "... could have a question measuring creativity because ASAT was oriented to academic subjects which involve an analytical and ordered mind, while drama or music leads to a more creative mind."

It was believed that ASAT only catered for the academic type subjects. Students believed that ASAT tested these subjects well and recognised that it ignored the creative aspects of education such as music, art, textiles, and photography. Following from this belief students considered ASAT was a valid indicator of likely university performance but was not particularly appropriate for measuring suitability for more practical courses or a job.

There were also some complaints that it did not test a person's commitment to their work.

> "a students application is not measured"
>
> "... results ignore application"
>
> "... it doesn't measure a person's stamina in doing assignments"

The general view was that ASAT was necessary as a moderating instrument but could only be considered as an indicator of ability in the traditional academic subjects. Students agreed that it could not test creativity but were divided about whether it tested humanities or the mathematics and science area most. The opinion expressed was heavily dependent upon whether the interviewee did humanities or mathematics and science. The following were two interesting points of view put forward by one student:

> "It is suitable as a moderator but individual marks do not deserve importance."
>
> "It doesn't matter that it is not suitable for art and music because they are not competing for the same positions."

G The Possibility of Sex Bias

It was particularly interesting to find that few students were aware of the sex bias issue. If they were aware of it they did not have strong views about it. The strongest comments came from the boys. Students said that ASAT reinforced sex roles expressed in society but did not believe that this would influence student performance.

> "It doesn't make any difference whether you're a female or a male as to how well you go on ASAT."

Only one student suggested that ASAT was biased against females, because of the way they thought. He could not expand or explain this view.

All the other students interviewed believed that if a bias existed it was caused by differential performances by students studying different subjects. A particularly interesting comment was made by one of the girls:

> "A lot of boys leave schools to do apprenticeships and stuff like that, but the lesser ability girls just stay on and do subjects like typing. They may pull down the average of the girls' marks."

The impression gained from the interviews, in general, was that the students could not identify any reason for sex bias in ASAT. They believed that it may contain subject bias, yet even this was not a consistent attitude among students.

Female Menstrual Cycles

Of the 14 girls interviewed, only one mentioned that she knew a girl who was menstruating on the day of testing and this may have affected her performance.

Queensland

The Queensland system of Tertiary Entrance Score calculation received relatively little public and community discussion, when compared to the concern expressed in the ACT. The Queensland and ACT systems were, however, similar in terms of the techniques used to assess students. Both used ASAT in conjunction with school-based assessments to calculate a TE score. In Queensland, however, unlike the ACT, students' personal ASAT scores were not reported. The interviews of Queensland students were carried out two to three months after students completed ASAT. It is recognised that this might have altered the perspective from which the students viewed ASAT, and the strength of their reactions to the questions asked.

A Students' understanding of the use of ASAT

In general, students in Queensland did not know the details of ASAT's use. All of the 30 students who were interviewed were aware that the mean of the school was very important because it was used to compare between schools.

> "All I know is that the school has to go well on ASAT so that my marks won't go down."

This comment by one student reflected the understanding of ten of the 30 students. The remaining 20 were also aware that the performance of the students in each of their subjects was important.

> "ASAT is used to put all the Queensland schools on a common scale. So that an A in one school means the same as an A at another school. Otherwise it wouldn't be fair because one teacher could be slack and give out high marks, easily where a hard teacher would give low marks. It's also used to compare between subjects but I don't know how they do it."

The above comment was representative of the general level of students' understanding about the use of ASAT.

It was particularly surprising to find that two students knew nothing about ASAT until a couple of weeks before they were scheduled to sit it.

> "I only heard that we had to do this ASAT test a week before we did it."

> "Nobody told me anything about an ASAT test until the middle of the year."

It is unlikely that these comments reflected typical levels of understanding because it appeared from the rest of the interview that both of these boys were somewhat uninterested in doing well.

Students in the interview sample had only a limited and basic knowledge about the use of ASAT. They knew it was used to moderate school assessments but they knew little about the statistical procedures employed. The lack of detailed knowledge did not appear to cause any major problems. Students who wanted a TE score were aware that their marks would be improved if they did as well as they possibly could. They understood that any extra marks that they were able to get by working hard would increase the school's average and this was to their own benefit.

B Differences in preparation and coaching between schools.

All of the students interviewed had experience doing examinations so most were prepared for ASAT. They were familiar with examination procedures and techniques. They all believed that they had sufficient experience with multiple choice items since they had done them in subject tests and examinations. However, it was clear that there were differences between the schools in the preparation undertaken for ASAT. Of the five non-Government schools visited, very intensive preparation was done at three. This included the use of trial ASAT tests. Although students did not know the source of the items used for the trial it was believed that they came from a previous ASAT which schools had managed to obtain.

Students at each of these schools were taken through the Students Information Bulletin discussing the answers to items.

> "We went through the booklet of questions and answers and discussed why the answers they picked were correct."

The Government schools had a less organized approach to ASAT preparation. None of the students from Government schools said that they went to sessions organized especially for practice at ASAT type items. It was left to subject teachers to provide this advice.

"We did a little bit of preparation in maths, but not a lot."

Some important hints had been given to all students.

"Yes, it was clearly stated that we should do every question."

"For any (question) we couldn't do teachers told us to pick a letter, like A, and just put it down the page."

The emphasis in the Government schools was on reading the Students Information Bulletin carefully. Students were told that further specific preparation would not be an advantage. However, while all the students interviewed had seen the Students Information Bulletin, only 19 had studied it carefully. The 11 students who didn't study it seemed not to care about ASAT.

"I didn't worry about it because I don't even want to go to Uni anyway."

Since ASAT was believed to be used only to obtain individual students' tertiary entrance scores, many of the students not interested in going on to further education did not worry about ASAT. One boy from a Government school said that if other students were not going to try, there was no point to his trying harder.

"I want to go to University but my ASAT can't help because some kids at this school don't try."

C Apprehension levels of students

As mentioned previously all of the students who were interviewed had experience with examinations and consequently did not approach ASAT with high levels of apprehension. One girl, who was not interviewed, approached one of the interviewers showing a degree of concern about ASAT. She was the only person encountered whose performance might have been affected by nerves and apprehension.

About half the students said that they were nervous prior to ASAT but they settled down once the exam started.

"I was a bit nervous before it but it wouldn't have affected how I went."

"I only had normal exam nerves."

"Everybody gets a little bit nervous before doing any sort of test."

The girls did not show greater apprehension than the boys.

Of the 15 males interviewed two said they always got quite nervous before exams, while three of the 15 girls said that were a little nervous. The striking difference between boys and girls was the number who were in a group that could be called "the I-don't-give-a-damn category". Of the 15 boys five said that they didn't even care about ASAT.

> "I didn't care about doing well."
>
> "ASAT isn't important if you don't want to go to Uni and stuff."

These attitudes were not found among the girls. They had a somewhat more serious approach to ASAT.

D The effect of course type on ASAT performance

Of the 30 students interviewed 18 could be classified as mathematics and science, 10 as humanities and two mixed. Reactions to the possibility of subject bias were mixed. Seven of the ten humanities students said that mathematics and science would be an advantage. However ten of the 18 mathematics and science students considered that extra English and humanities would have helped them with ASAT. One of the students doing a mixed course saw English as an advantage because of the level of sophistication of the language used.

> "The passages and instructions in ASAT are complicated to understand. If you are good at English it is a big help to do the questions well."

The girls' views on subject advantage did not differ from those of the boys. The seven humanities students who saw mathematics and science as an advantage consisted of four girls and three boys. The ten mathematics and science students who saw humanities as an advantage comprised six boys and four girls.

When students commented that one subject area gave some candidates an advantage, they had difficulty providing a reason to justify their statement. One of the mathematics and science boys said:

> "I am not that good at drama and clear thinking so if I did more English I would have done better, I reckon."

Another boy who thought mathematics and science would be an advantage said:

> "The biology and mathematics stuff was the hardest so I would have done better if I studied more of those subjects."

The students who said that some subjects would be an advantage did so on the basis of their own weaknesses. They believed that students who were studying the subjects that they themselves were not good at had a clear advantage. They also said that ASAT contained more items from the subjects that they were poor at, or were not studying.

> "There was a lot more comprehension than I expected."
>
> "There were too many maths questions."

As a group, the students did not provide any argument that suggested that the style or the format of the test suited students with any particular background.

E Strategies employed in working through the test.

All the students interviewed had examination experience and this was reflected in the strategies they employed when doing ASAT. Moreover, all of the students interviewed allocated time to each question. Eighteen started at the beginning and went straight through. Of those who did this, most omitted difficult questions and later returned to them. Others guessed answers as their time allocation for the item expired.

The remaining 12 students adopted one of the following strategies:

(a) they did the sections that looked easiest first, because these came from their area of study or because they did not involve a lot of reading; and

(b) they did the harder units and those with a lot of reading, to make sure that they got them done.

It was clear from the interviews that students had used methods that would enable them to perform best on ASAT.

F Validity of ASAT as perceived by students.

The students interviewed all saw ASAT as important in 'levelling' the schools.

> "ASAT is necessary because different schools may give high marks more than other schools. With ASAT you can check that this doesn't happen."

ASAT's role as a moderator between schools was seen as necessary and valid. It was seen as providing students from all schools, regardless of standard, with an even chance of obtaining an adequate TE score.

The opinion was expressed, however, that students who did not want to go to University and did not want a TE score might not try as hard to do well. If a school had more of these students than another school then the mean would be lower at that school. This was seen as not being fair to the students who did want a TE score.

At an individual level, most students did not see an ASAT score as a valid personal indicator for university entrance.

> "ASAT ignores too much of what you learn at school."
>
> "It doesn't cover a wide enough range of the subjects I do."

In their criticism it was clear that students did not understand the meaning of 'scholastic aptitude'. They criticized ASAT's suitability on content grounds rather than on its ability to test the skills necessary to succeed at University. Students did not consider that a personal ASAT score would have any relevance for a job.

G Possibility of sex bias

No thought of sex bias was found among the students interviewed. They had not seen any mention of it until they did the questionnaire and at that time wondered why it was an issue.

> "I didn't think anything about it until the survey we had to do. Now that was biased and sexist."

None of the students could identify a way or ways in which ASAT might contain a bias against females.

Female menstrual cycles

Two girls mentioned they could see performance being affected if a girl was menstruating at the time of an examination. None of the students actually had problems themselves.

Summary and Conclusions

The students from the two systems had quite different ideas about ASAT. In the ACT considerable press coverage of the arguments surrounding the use of ASAT had heightened students' awareness of many issues the interviews were designed to investigate. Students in the ACT had thought more carefully through arguments about the tertiary entrance system. The Queensland students had accepted ASAT as part of their tertiary selection procedures, and with less public debate about ASAT's use, students had not thought through the issues involved.

In Queensland, fewer students had knowledge of the procedures involved in ASAT's use but they all had a basic level of understanding regarding its purpose. In the ACT more students were familiar with the moderation procedure. However, there were large variations in the students' level of understanding. Some students knew nothing, while others had a clear understanding of the statistical procedure and the possible effects ASAT could have on their school assessments.

Owing to the greater public controversy over ASAT in the ACT students wanted to know how it was used. Those who didn't know how it was used were unnecessarily worried about its effect on their marks. They tended to be confused about its importance. Queensland students accepted it as an examination on which they had to perform well if they hoped to improve their chances of entering university. They knew it was only one component of their TE requirements which were influenced more significantly by the marks in their coursework.

In both Queensland and the ACT there were large differences between schools in the amount of preparation and coaching available to students. Non-government schools provided more well attended preparation sessions than did the Government schools. All schools made the Students Information Bulletin available and provided students with basic information about test length, format and question type. In the ACT students said that they sat the trial ASAT in examination conditions. There was however, disagreement among the students about the usefulness of the Students Information Bulletin and the trial ASAT as preparation. Only one school had explained ASAT's use in a way that enabled students to understand clearly the possible effects on school assessments. This helped the students in their approach to the examination. It should also be noted that the Sample Collection of ASAT questions is distributed in the ACT, but not to Queensland students.

In both the ACT and Queensland, few of the students expressed that doing ASAT had especially upset them. When girls described how nervous they were it did not indicate that they were more apprehensive than the boys. However, it is difficult to generalize about this since the students who were interviewed were generally considered to be of above average ability. In Queensland, students accepted ASAT as a small, but important, part of their tertiary entrance requirements and did not show unnatural levels of anxiety. The public debate on the validity of ASAT in the ACT might have resulted in students there being more apprehensive about ASAT's effect on their marks.

The issue of mathematics and science students being advantaged by ASAT arose in the ACT but did not appear to arise in Queensland. Queensland students generally saw ASAT as favouring other students who took those subjects that they personally did not study or found difficult. In the ACT, girls said that mathematics was an advantage. The ACT boys, like the students in Queensland, said that doing more of the subjects that they were not familiar with would have been of the greatest advantage.

Students in Queensland had greater experience in examinations and this was reflected in their method of doing ASAT. All of the Queensland students used some method or strategy, whereas six of the 29 ACT students just started at the beginning and worked straight through using no strategy or time allocation system. Nine of the ACT students used complex strategies in doing ASAT while 14 of the Queensland students used complex strategies. Complex strategies were not used more often by students from any subgroup. In the ACT all mathematics and science students used some type of strategy and this might have been a result of their greater exposure to tests.

In both the ACT and Queensland, students did not believe that an individual's ASAT mark deserved importance. This was a greater issue in the ACT because a student's ASAT mark appeared on the student's profile of information. Some students said that ASAT was no good because it didn't test the content of subjects that they had studied. This arose because students did not understand that ASAT was a test of 'scholastic aptitude', defined as:

> "... a capacity to undertake intellectual pursuits common to scholarship at senior secondary and post-secondary levels." (ASAT test specification)

This meaning of 'scholastic aptitude' appeared not to have been explained to students.

Comments expressing doubt about ASAT's suitability for students in the non-academic subjects such as art and typing were only made in the ACT. The Queensland students who were interviewed had not thought about the issue until prompted by the interviewer. Few students referred to the suitability of ASAT as a moderator of assessment or performance in the modern languages. This was probably a consequence of the small number of students who studied these subjects at Year 12.

There were some students who expressed doubts about the accuracy of ASAT scores on other grounds. It was expressed by more than one student (from both Government and non-Government schools) that cheating had occurred.

> "I know this doesn't go on at our school but I have heard that some schools cheat."

Students claimed that schools had cheated the system in a variety of ways. First, the claim was made that schools actively encouraged less able students to stay away on the days of testing. Secondly, there were claims that inadequate supervision was provided by school teachers. None of the students admitted that these activities occurred at their own school. Students had heard rumours of cheating taking place rather than had first-hand knowledge. For this reason it was difficult to establish the accuracy of such comments. One student from a Catholic boys school said that:

> "In the end someone has to be trusted and the professionalism of teachers should be trusted."

School-based assessment was generally seen as better than external examinations and it was recognized that a measure of some kind was necessary to moderate between schools and subjects. The students that were interviewed showed a sophistication in their knowledge about assessment that may not have been expected two decades ago. ACT students had mixed feelings about ASAT's usefulness but were not able to offer alternatives, within the framework of school-based assessment. Queensland students saw that ASAT had its problems but generally regarded it as necessary and valid. Some Queensland students believed that ASAT was not a valid indicator of suitability for university entrance; however, they were more concerned about the fact that less capable students, by not trying hard on ASAT, may lower their TE score by decreasing their school's ASAT mean.

Of the 29 females interviewed none said they had any physical or health problems at the time of testing. Three said that a girl may not be able to work to her best ability if she was menstruating at the time of the examination.

Few students in the ACT and none in Queensland were aware of the sex bias issue. Some students in the ACT said that it contained a degree of content bias through reinforcement of sex role stereotypes, but it was not suggested that this would affect performance. Some of the ACT students who were aware that males outperformed females on ASAT believed it was a result of differential coursework, while others could not identify reasons for differences in performance.

CHAPTER 6

STUDENT ATTITUDES TOWARDS ASAT

It has been generally accepted that males outperform females at higher level mathematics and females perform better than males in verbal areas (Maccoby and Jacklin, 1974). However, more recent evidence has indicated that, if various other factors were taken into account, there was uncertainty about the significance and size of sex differences. Recent studies (Fennema and Sherman, 1977; Pallas and Alexander, 1983) have found that sex differences were largely eliminated if experience in a subject area had been controlled. Fennema and Sherman (1977) hypothesized that both cognitive variables and affective variables were associated with sex related differences in mathematics achievement. They found few sex related cognitive differences but many sex related attitudinal differences. Since ASAT is a test that measures, in part, mathematical aptitude, it was expected that sex differences associated with attitudinal variables could be directly related to sex differences in performance on ASAT.

Fennema and Sherman (1977) used eight affective measures in a study of sex differences in attitudes towards mathematics, which have become known as the Fennema-Sherman Attitude Scales (Fennema and Sherman, 1976). The eight variables tested by the scales were: Confidence in Learning Mathematics, Effectance Motivation, Mathematics as a Male Domain, Attitude Towards Success in Mathematics, Usefulness of Mathematics, Mother, Father, Teacher. The last three scales were designed to measure perceived attitude of important others towards the individual as a learner of mathematics. The term Effectance Motivation was used by Fennema and Sherman to refer to the enjoyment students gained from the study of mathematics and the stimulation they gained from it. Other research studies (Milton, 1957; Carey, 1958; Husén, 1967; Neale, 1969; Aiken, 1972, 1974; Hilton and Berglund, 1974) have supported the hypothesis that a student's attitudes towards a subject play an important part in performance in that subject.

It was considered that the evidence available on attitudes towards mathematics necessitated the undertaking of an investigation into the possibility of sex differences in attitudes towards ASAT, and the effect that these differences might have on performance on ASAT.

Scale Development

Four Fennema-Sherman scales were adapted for use in testing attitudes in relation to ASAT, namely: Confidence in Learning Mathematics, Mathematics as a Male Domain, Attitude towards Success in Mathematics, and Effectance Motivation. After adaptation, these scales became known as: Confidence in Success on ASAT, ASAT as a Male Domain, Attitude towards Success on ASAT, and Effectance Motivation.

The Fennema-Sherman scales each consisted of twelve items. Six of these were positive and six were negative. Each of the Fennema-Sherman items on the above four scales were rewritten so that they focused on ASAT rather than mathematics. Approximately 50 per cent of the items could be altered by replacing the word 'math' with 'ASAT'. The remaining 50 per cent were re-worded so that 'math' could be replaced by 'ASAT'. Additional items were written and included in order to increase the size of the available item pool.

The items were examined critically by a panel consisting of experts in test development, item writing and attitudes to mathematics. After examination, 28 items remained, seven for each scale. It was decided that a five-point rating scale response type should be used. Half of the 28 items had positive wording and half had negative wording. The 28 items were randomly mixed and placed at the end of the ASAT General Information Questionnaire. This questionnaire is presented as Appendix 6.1.

The attitudes that the scales sought to measure may be described as follows:

Scale 1 Confidence in Success

This scale ranged from a confident student feeling about ASAT:
'I felt I could perform well on ASAT',
to a strongly negative attitude:
'I am not the type to do well on ASAT'.

Scale 2 Effectance Motivation

This involved students' feelings regarding the challenge involved in the solution of ASAT items. The scale ranged from:
'The ASAT questions were enjoyable and stimulating to me',
to:
'Sitting for ASAT was boring'.

Scale 3 ASAT as a Male Domain

This scale sought to ascertain whether the potential for success on ASAT had a masculine orientation. Items ranged from:
'Females can do just as well as males on ASAT'
to:
'It's hard to believe a female could do very well on ASAT'.

Scale 4 Attitude towards Success

The Attitude towards Success scale sought to measure a student's reactions to performing well on ASAT. Items ranged from:
'I'd be proud to be an outstanding student on ASAT'
to:
'I wouldn't like people to think I was very good at ASAT'.

The Samples

The questionnaire was distributed to three samples of students from Queensland, Western Australia and the Australian Capital Territory. From a total sample of 1780, completed questionnaires were returned by 1408 students giving a response rate of 79 per cent. Table 6.1 shows the distribution of returns from the three States.

In the Australian Capital Territory (ACT), 50 questionnaires were sent to each of the 13 colleges and only 10 were sent to School Without Walls (SWOW), since SWOW did not have as many students as the colleges. In Queensland, 10 questionnaires were sent to each of 52 schools. These schools were selected at random from a list of all the schools that provided Year 12 courses in the Brisbane area. In Western Australia, 20 questionnaires were sent to each of 30 schools selected at random with probability proportional to size. Teachers were asked to administer the questionnaire to students selected at random from the school. In Western Australia and Queensland this worked well; however, in the ACT some teachers selected entire class groups for the administration of the questionnaire.

In the ACT, questionnaires were administered the day after students sat for ASAT. In Western Australia, students were asked to complete the questionnaire during the period while they were sitting for their public examinations in November, 1983. During this period they had taken the ASAT test. Queensland students did the questionnaire about two months after sitting for ASAT.

Table 6.1 Size and Sex Composition of Samples

	Australian Capital Territory	Queensland	Western Australia	Total (%)
Designed sample	660	520	600	1780
Achieved sample:				
Male	251	224	217	692
Female	269	237	209	715
Total (%)	520 (79)	461 (89)	426 (71)	1407[a] (79)

[a] One student did not indicate sex and had to be omitted from these analyses.

Factor Analyses

Item responses were recorded so that all items were coded in a positive direction (i.e. one and two-an unfavourable reaction, four and five-favourable and three-uncertain). Missing data were allocated to response category three indicating uncertainty. This had little effect on the analyses since the inclusion of missing data on particular items had no significant bearing on the results.

A principal components factor analysis was carried out using unity as the initial communality estimate. Six factors were found to have eigenvalues greater than one. However, only four factors were retained after applying a scree test, as was consistent with the design of the instrument. These four factors were then rotated using a varimax rotation. Factor loadings greater than 0.30 were recorded (Nunnally, 1978) and these factor loadings have been reported in Table 6.2. The four factors were identified as Confidence in Success, Effectance Motivation, ASAT as a Male Domain and Attitude towards Success on ASAT. Items 15 and 25 were omitted because they did not load on the attitude scales to which they were assigned.

The same procedure was carried out for each sex within each State. Results of these analyses are recorded in Appendix 6.2. The same four factors were obtained on each occasion. The first three factors, Confidence in Success on ASAT, Effectance Motivation, and ASAT as a Male Domain, held together well but the fourth factor, Attitude towards Success on ASAT, was not as clear. It should be noted that for the sex within State analyses there were insufficient cases per variable for a confident factor analysis.

Table 6.2 Factor Patterns of Attitude Scale Items[a]

Item	Confidence in Success	Effectance Motivation	ASAT as a Male Domain	Attitude towards Success on ASAT
3	69			
6	69			
7	74			
11	77			
12	70			
22	70			
27	69			
1		70		
2		45		
9		74		
17		77		
20		76		
24		77		
4			61	
8			70	
10			72	
18			76	
21			62	
26			72	
5				62
13				47
14				41
16				54
19				67
23				60
28				39
15				
25			36	33
Percentage variance	14.4	12.9	11.4	8.4

[a] Loadings >0.30 are shown. All factor loadings are multiplied by 100. N = 1407.

Table 6.3 Reliabilities for Attitude Scales

Scale	Confidence in Success	Effectance Motivation	ASAT as a Male Domain	Attitude towards Success on ASAT
Number of items	7	6	6	7
Alpha reliability coefficient	0.87	0.84	0.78	0.62

Nunnally (1973) recommended ten cases per variable and this criterion was not satisfied for these analyses.

The coefficient alpha reliabilities (Cronbach, 1951) were then calculated for each scale. They are shown in Table 6.3. Items were only included in a scale if they had a factor loading of greater than 0.30 on that scale and they added to the reliability of the scale.

Results

Figure 6.1 shows profiles of students' scores for each State and each sex on the four scales. The scores on each scale were standardized to a mean of 0 and standard deviation of 1. Figure 6.1(i) shows that the Queensland students had more favourable attitudes than the students from the ACT and Western Australia for three of the scales. Queensland students had higher confidence in success, greater effectance motivation, and they saw ASAT as equally suited to male and female students. Figure 6.1(ii) records the sex differences on each of the scales. On the Confidence in Success scale, males

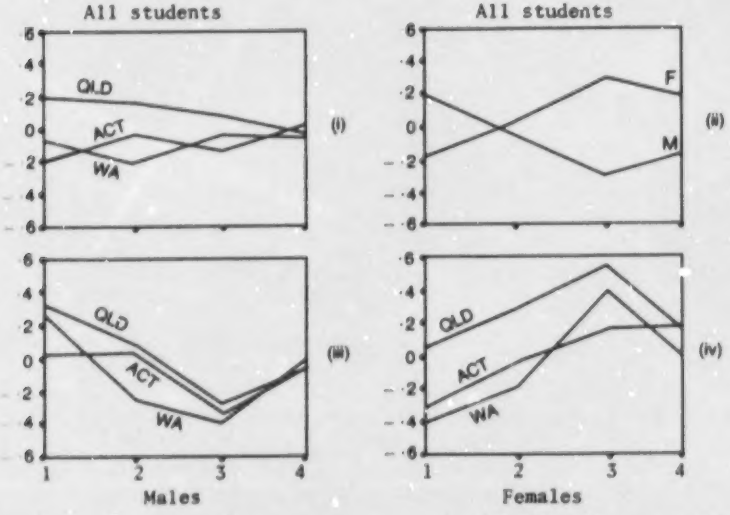

Figure 6.1 Profiles of Student Scores.

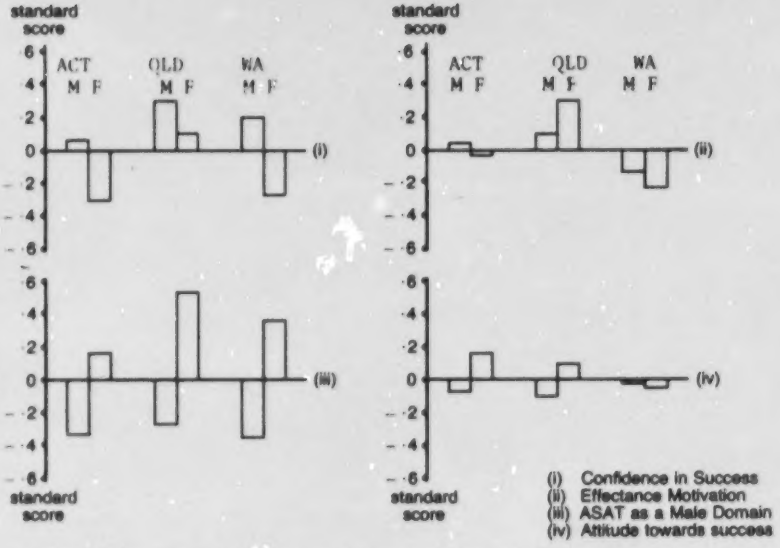

Figure 6.2 Comparison of Male and Female Scores.

scored more highly than females while on each of the other scales females expressed the more favourable attitudes. Figure 6.1(iii) shows the relative scores of males over the three States. The ACT boys had the lowest confidence in success, while the Queensland and the Western Australian males were close together. Western Australian boys had the lowest scores on the Effectance Motivation scale while the boys from the three States scored close together on the other two scales. Figure 6.1(iv) shows the relative scores of the females from each State. The girls in Queensland had more favourable attitudes on the first three scales. For the Attitude towards Success scale the scores of the girls in Queensland and the ACT were close together.

Figure 6.2 compares the sexes for each of the scales and for each State. Figure 6.2(i) shows that in both the ACT and Western Australia, females had lower than average confidence in success. There was a large difference between the males and females in Western Australia. Figure 6.2(ii) shows the lower effectance motivation of Western Australian students to be consistent for both sexes. The ACT males scored slightly higher on this scale than the females. Figure 6.2(iii) records consistently more favourable attitudes for girls than boys on the ASAT as a Male Domain scale indicating that the girls did not see ASAT as a male domain. The differences on the Attitude towards Success scale as reported in Figure 6.2(iv) were small.

The question of whether or not the differences between the sexes and the States presented in Figure 6.1 and Figure 6.2 were significant was a complex one, and analysis of variance was used to explore the differences.

Table 6.4 Attitude Scale Statistics for each State

Group		N	Schls	Confidence in Success on ASAT				Effectance Motivation				ASAT as a Male Domain				Attitude Towards Success			
				$\bar{x}$	std	sk	krt	$\bar{x}$	std	sk	krt	$\bar{x}$	std	sk	krt	$\bar{x}$	std	sk	krt
ACT	Males	251	11	23.81	5.39	-0.32	-0.60	18.65	5.59	0.02	-0.99	24.23	4.13	-0.42	-0.32	28.00	3.50	-0.43	0.40
	Females	269	12	22.07	5.53	-0.20	-0.56	18.37	5.13	-0.11	-0.71	26.26	3.49	-0.76	-0.32	28.96	3.59	-0.53	0.21
	Total	520	14	22.91	5.52	-0.25	-0.59	18.51	5.36	-0.03	-0.85	25.28	3.94	-0.63	-0.24	28.50	3.58	-0.46	0.24
Qld	Males	224	42	25.53	4.59	-0.52	0.47	19.02	5.25	-0.15	-0.75	24.49	4.13	-1.05	1.94	27.92	3.81	-0.71	1.76
	Females	237	38	24.23	4.90	-0.47	-0.16	20.22	5.11	-0.47	-0.59	27.79	2.59	-1.08	0.44	28.71	3.13	-0.35	0.16
	Total	461	52	24.86	4.79	-0.50	0.11	19.63	5.21	-0.30	-0.73	26.19	3.80	-1.29	2.54	28.32	3.50	-0.63	1.46
WA	Males	217	29	24.96	4.95	-0.65	0.87	17.64	5.76	-0.11	-0.73	24.07	4.53	-1.33	2.81	28.22	3.42	-0.59	0.77
	Females	209	20	21.96	5.37	0.00	-0.60	17.27	5.21	0.10	-0.81	27.04	3.33	-1.02	0.24	28.23	3.06	-0.07	-0.40
	Total	426	33	23.49	5.37	-0.32	-0.26	17.46	5.49	-0.01	-0.76	25.53	4.25	-1.33	2.81	28.23	3.24	-0.38	0.35
All States	Males	692	82	24.73	5.05	-0.51	0.09	18.46	5.56	-0.09	-0.80	24.27	4.25	-0.95	1.59	28.04	3.58	-0.59	1.07
	Females	715	70	22.76	5.37	-0.25	-0.53	18.66	5.28	-0.15	-0.83	27.00	3.23	-0.99	0.22	28.67	3.30	-0.33	0.03
	Total	1407	99	23.73	5.31	-0.38	-0.31	18.56	5.42	-0.12	-0.81	25.65	4.01	-1.07	1.61	28.36	3.45	-0.49	0.69
Number of Items in Scale				7				6				6				7			

Analysis of Variance

Assumptions

One of the assumptions associated with an analysis of variance is that the dependent variable is normally distributed in the population from which the sample is drawn. It was considered that this would be the case with the four dependent variables used in these analyses. A second assumption is that the variances of the scores on the variables measured for each of the sub-populations were equal. This is the homogeneity of variance assumption. Table 6.4 reports summary statistic for the distributions of the dependent variables in the samples drawn. From the statistics in Table 6.4, it can be seen that the variances were similar. Most of the distributions were slightly negatively skewed and a little flatter than the normal distributions, as indicated by the negative kurtosis values.

Although the data used in the analyses which follow did not have exact normal distributions, it was likely that the conclusions drawn from the data using an F test would not be seriously affected, because:

> One advantage of the Analysis of Variance is that reasonable departures from the assumptions of normality and homogeneity may occur without seriously affecting the validity of the inferences drawn from the data. (Ferguson, 1981:246)

Design Effect

The F values and the significance levels reported from an analysis of variance are based on the independence of observations associated with simple random sampling. However, on most occasions, when sampling is undertaken in educational research, the samples that are drawn are, in actual fact, some type of cluster sample. That is, schools are generally selected at random with probability proportional to size, while students within the schools are clustered in classes. The use of significance tests based on simple random sampling methods, when other sampling techniques have been used, may lead to erroneous conclusions. Kish (1965) has argued that the use of simple random sample formulae on data from complex samples is the most frequent source of gross mistakes in the construction of confidence intervals and tests of hypotheses in educational and social science research.

Various studies have investigated the effect that sample design may have on the significance tests provided for most statistical measures. In an empirical study in Australia, Ross (1976) found that the design effect for means, when clustering occurred within schools, was substantial. He estimated

Table 6.5 Analysis of Variance Results for Confidence in Success on ASAT

	F Ratio	Corrected F Ratio	df	Significance Simple	Corrected
Sex	53.56	13.39	1	$p < 0.001$	$p < 0.001$
State	18.12	4.53	2	$p < 0.001$	$p < 0.05$
Sex X State	3.25	0.81	2	$p < 0.05$	ns

	Sex		
State	Male	Female	Total
Australian Capital Territory	23.81	22.07	22.91
Queensland	25.53	24.23	24.86
Western Australia	24.97	21.96	23.50
Total	24.73	22.76	23.73

that the design effect was such that F values should be divided by the value four. The sampling in this study was similar to one of the methods of cluster sampling explored by Ross (1976) and consequently the F value in the following analyses of variance tables have been divided by the estimated value of four in order to obtain a more accurate indication of the significance of the differences in means.

Scale 1 Confidence in Success on ASAT

Table 6.5 presents the results for the analysis of variance of the scores on the Confidence in Success scale, when examined with respect to sex and State.

Table 6.5 shows that there was a significant sex difference in scores on the Confidence in Success scale. As in indicated in Table 6.5, the males scored higher than the females on this scale. There was also a significant State difference. The Queensland students scored higher than both the ACT and Western Australian students. There was no significant interaction between State and sex after design effects were taken into account.

Scale 2 Effectance Motivation

The analysis of variance results for the Effectance Motivation scale are reported in Table 6.6 and this table also shows the mean scores for males and females and for each of the States. There were no significant sex differences while the State difference was significant after design effects were taken into account. There was also no significant interaction between sex and State after design effects were considered. The scores for the

Table 6.6 Analysis of Variance Results for Effectance Motivation

	F Ratio	Corrected F Ratio	df	Significance Simple	Significance Corrected
Sex	0.52	0.13	1	ns	ns
State	18.17	4.54	2	$p < 0.001$	$p < 0.01$
Sex X State	3.14	0.78	2	$p < 0.05$	ns

	Sex		
State	Male	Female	Total
Australian Capital Territory	18.65	18.37	18.51
Queensland	19.02	20.21	19.64
Western Australia	17.64	17.27	17.46
Total	18.46	18.66	18.56

Queensland sample of students were high, and that of the Western Australian sample were low on the Effectance Motivation scale. The sample of students in the Australian Capital Territory recorded scores approximately midway between the two other state samples.

Scale 3 ASAT as a Male Domain

Table 6.7 reports the results of an analysis of variance for the ASAT as a Male Domain scale. The results show that there was a significant sex difference for this scale. Table 6.7 also shows that the females scored

Table 6.7 Analysis of Variance Results for ASAT as a Male Domain

	F Ratio	Corrected F Ratio	df	Significance Simple	Significance Corrected
Sex	190.54	47.63	1	$p < 0.001$	$p < 0.001$
State	7.17	1.79	2	$p < 0.001$	ns
Sex X State	3.82	0.95	2	$p < 0.05$	ns

	Sex		
	Male	Female	Total
Australian Capital Territory	24.23	26.26	25.28
Queensland	24.50	27.79	26.19
Western Australia	24.07	27.04	25.53
Total	24.27	27.00	25.65

Table 6.8 Analysis of Variance Results for Attitude Towards Success on ASAT

	F Ratio	Corrected F Ratio	df	Significance Simple	Significance Corrected
Sex	10.35	2.59	1	$p < 0.01$	ns
State	0.67	0.17	2	ns	ns
Sex X State	2.42	0.61	2	ns	ns

	Sex		
State	Male	Female	Total
Australian Capital Territory	28.00	28.96	28.50
Queensland	27.92	28.71	28.33
Western Australia	28.22	28.24	28.23
Total	28.04	28.67	28.36

significantly higher on this scale, indicating that males considered ASAT a male domain more than did females. In Table 6.7, the F ratio for States indicated a significant State difference. However, this was not significant if design effects were taken into account. The interaction between State and sex was also not significant after taking design effects into account.

Scale 4 Attitude Towards Success on ASAT

Table 6.8 shows that no significant sex, State or interaction effects were found on the Attitude towards Success scale after design effects were considered. Although the females had a slightly more favourable attitude towards success as indicated in Table 6.8, this difference did not reach significance.

Discussion

Male students significantly outscored females on the Confidence in Success scale. The males showed a greater confidence in success than the females, and in addition this result was consistent over the States. This finding was in agreement with the body of evidence concerning sex role stereotyping and socialization effects within society. Fennema and Sherman (1977) found significant sex differences in confidence in success in learning mathematics. They argued that there was a strong link between confidence in success and performance. The argument that they have presented could be extended to confidence and performance on ASAT.

> It might be argued that girls are not so confident in mathematics because they achieve more poorly than males. This seems unlikely since: (a) the sex related differences in confidence occur before sex related differences in mathematics performance are generally observed; (b) there are more instances of sex related differences in confidence than in mathematical achievement and when there is a sex related difference in achievement, there is also a sex related difference in confidence, but not always vice versa; (c) the lesser confidence of girls in their ability to perform in mathematics is consistent with their lesser confidence generally.... (Fennema and Sherman, 1977:67)

Significant correlations between attitude and performance have been found in a number of studies (Anttonen) 1967; Husén, 1967; Ryan, 1968; Rounds and Hendel, 1980) but a causal link has not been established. Neale (1969) argued that it was not possible to establish a causal link between attitude and performance without using an experimental study that manipulated students' attitudes.

On the Effectance Motivation scale there were no significant sex differences. This was consistent with the Fennema and Sherman (1977) study which did not find significant sex differences. Rounds and Hendel (1980) found that Fennema and Sherman's Effectance Motivation scale correlated positively with an arithmetic achievement test. Although no causal link could be established, it may have been that the Queensland students were gaining some benefit from their higher scores on this scale.

The ASAT as a Male Domain scale showed a clear sex difference. The boys considered ASAT as a male domain to a much greater extent than did the girls. The females saw themselves to be just as suited to ASAT as the males. However, the males did not share this view. This result was particularly interesting and must be considered in one sense contradictory to the results obtained on the Confidence in Success scale. The girls seemed to be saying on the one hand that ASAT was not a male domain, that is, that females had the ability to perform just as well on the test as males. On the other hand, their lower score on the Confidence in Sucess scale seemed to contradict this. There are two possible bases for this apparent contradiction. First, by obtaining high scores on the scale ASAT as a Male Domain, the girls were saying that females, in general, could score as well as males on ASAT. By obtaining a significantly lower score on the Confidence in Success scale the girls were saying that they, as individuals, did not have confidence in success. Secondly, the girls responses to the ASAT as a Male Domain scale might have been a reaction to the somewhat chauvinistic implication that males could do better than females at anything.

There were no significant sex differences on the Attitude towards Success scale after considering design effects. The trend was for females to outscore the males but the difference did not reach significance. This was an unexpected result because previous research had indicated that females expressed what might be called a "fear of success" in areas not seen as suitable for them (Horner, 1972; Stein and Bailey, 1973; Leder, 1979).

The results, as displayed in Figure 6.1, showed that the Queensland students scored consistently higher on both the Confidence in Success and Effectance Motivation scales. Perhaps the lower scores in the ACT could, to some extent, be explained by the fact that ASAT has been subjected to open criticism and public discussion within the ACT. This criticism and its effect on students was apparent in the evidence from the interviews presented in the previous chapter. Students' attitudes towards ASAT could well have been affected by the public debate over ASAT's validity. The lower scores obtained on the three scales by the ACT students might have been partially due to the questioning of the test's validity. The heightened awareness, public debate and concern over ASAT that was evident in the ACT interviews were not apparent in the interviews with the Queensland students where ASAT had not come under such scrutiny. This public debate and controversy might have affected the attitudes of students as they approached ASAT. It should be noted that interviews were not carried out in Western Australia and comparison can only be made between the ACT and Queensland on this aspect of the study.

It should also be noted that the students from Queensland responded to the attitude scales some time after they sat for ASAT. The ACT and Western Australian students responded to the attitude scales very close to the time they did the ASAT test. There was, therefore, the possibility that the Queensland students had forgotten, to some extent, about the anxiety and pressure and that this created more favourable attitudes in comparison with the other States.

Summary and Conclusions

A clear sex difference was found on the scales Confidence in Success and ASAT as a Male Domain. The males in each of the three States were more confident than the females in doing well on ASAT. This was consistent with previous research on confidence and performance in mathematics. In all three States, however, females considered that ASAT was not a male domain.

The contradictory nature of this result has been discussed. It was pointed out that these results might have been due to the individual nature of one scale (i.e. Confidence in Success) and the collective and perhaps reactionary nature of the ASAT as a Male Domain scale.

A significant difference between States was found for the Confidence in Success and Effectance Motivation scales. Queensland students consistently scored higher on these scales, compared to their ACT and Western Australian counterparts. This trend was also evident for the ASAT as a Male Domain scale. These differences were discussed in terms of the heightened awareness, criticisms, public debate and media concern over the validity of ASAT present in the ACT. It was noted that these issues were absent in Queensland and Western Australia. It was also believed that the timing of the questionnaire in different States might have had some effect on the students' responses.

Despite the fact that no significant sex or State differences were found for the scale Attitude towards Success, it was interesting to note that females tended to score higher on this scale than did the males. Finally, the importance of attitudes towards ASAT, as expressed by the scores on the four scales, was seen to be clarified and supported by the interview data presented in the previous chapter.

However, the findings recorded in this chapter concerning the differences between the sexes and between students in the three different state systems in their attitudes towards ASAT, while of interest, are not of themselves important for this investigation. What is important is whether the differences between the sexes in attitudes might be considered to contribute to differences in performance between the sexes on ASAT. In Chapter 9, the contribution of attitudinal differences to performance is examined, and an attempt is made to estimate the magnitude of the effects of sex differences in attitude on the observed differences between the sexes in performance.

CHAPTER 7

TIME SPENT IN PREPARATION

The evidence from previous research studies undertaken in the 1950s and 1960s suggested that the benefits of coaching for aptitude tests such as ASAT were of little practical significance (Dyer, 1953; French, 1955; Dear, 1958; Frankel, 1960; Whitla, 1962). However, more recent research has indicated that this view could no longer be held with confidence. (Pike, 1979; Slack and Porter, 1980, Messick and Jungeblut; 1980; Bangert-Drowns et al, 1983). There has also been considerable argument about the amount of preparation that would be the most effective. Messick and Jungeblut (1980) found a positive correlation between Scholastic Aptitude Test (SAT) scores and the time spent on preparation. This indicated that higher results on the SAT could be achieved by students if they undertook longer preparation programs. Bangert-Drowns et al (1983) suggested that a brief program would produce the best results, and their suggestion was supported by the unpublished findings of Kulik, Bangert and Kulik (1982) in which the effects of coaching programs on aptitude tests were examined.

The major aim of this chapter is to explore some issues of importance that were raised in both discussions with the authorities that used ASAT, and interviews carried out with students from the ACT and Queensland. The analyses were exploratory in so far as no hypotheses were postulated and therefore no explanations of the results were undertaken. It is hoped that the results will provide descriptive information and impetus for further research in the area.

The controversy raised by previous research studies identified some issues that this investigation could explore. In particular, they suggested that consideration should be given to the differences in the amount of preparation across States, between sexes and between school types (Government and non-Government). This study was able to investigate preparation in two ways. The interviews reported in Chapter 5 were used to ascertain whether students considered that they had had sufficient preparation for ASAT. The questionnaire (Appendix 6.1) on the other hand, examined the actual amount of time students spent in specific preparation for ASAT by posing the following questions:

"Were there any formal preparation sessions for ASAT at school? If yes, how many did you attend?"

"Did you spend time preparing for ASAT at home? If yes, approximately how many hours?"

Using the data from responses to these questions it was possible to explore differences between the sexes, States and school type in the amount of preparation undertaken by students. No attempt was made to relate preparation to performance as it was considered that the quality of the preparation could not be assessed and quality was perhaps the most important aspect of any preparation program.

The use of the questionnaire data on preparation enabled three questions to be studied:

1 Were there differences between the sexes in the amount of preparation undertaken by students?

2 Were there differences between the States in the amount of preparation undertaken by studnets?

3 Were there differences between Government and non-Government schools in the amount of preparation undertaken by students?

Results

In each of the analyses, levels of statistical significance are presented for simple random samples. The corrected values were estimated after consideration of design effects as discussed in Chapter 6.

School Preparation Sessions

The results of an analysis of variance for school preparation sessions with respect to sex and State are reported in Table 7.1. Females in the sample attended more sessions than the males, although this was not consistent over all States. In both the ACT and Queensland females attended more sessions than males while in Western Australia male students attended a greater number of sessions than did females. However, Western Australian students attended fewer sessions than the students from the ACT and Queensland.

The results obtained by analysis of variance showed that there were significant sex differences, significant State differences and a significant interaction. The significant interaction was probably obtained because males did more preparation than females in Western Australia, while females did more preparation than males in the ACT and Queensland.

Table 7.1 Analysis of Variance Results for School Preparation Sessions

	F Ratio	Corrected F Ratio	df	Significance Simple	Significance Corrected
Sex	18.67	4.67	1	$p < 0.001$	$p < 0.05$
State	113.86	28.46	2	$p < 0.001$	$p < 0.001$
Sex X State	24.93	6.23	2	$p < 0.001$	$p < 0.01$

R^2 = 19%

State	Male	Female	Total
Australian Capital Territory	1.18	2.55	1.88
Queensland	2.04	2.22	2.14
Western Australia	0.45	0.19	0.32
Total	1.27	1.84	1.56

Home Preparation Time

Table 7.2 reports the results for an analysis of variance of home preparation in hours with respect to sex and State. After design effects had been taken into account no significant differences were found in home preparation time. This was reflected in the particularly low R^2 value of 0.8 per cent.

Table 7.2 Analysis of Variance Results for Home Preparation Time[a]

	F Ratio	Corrected F Ratio	df	Significance Simple	Significance Corrected
Sex	0.55	0.14	1	ns	ns
State	1.56	0.39	2	ns	ns
Sex X State	3.99	1.00	2	$p < 0.05$	ns

R^2 = 0.8%

State	Male	Female	Total
Australian Capital Territory	1.47	0.93	1.19
Queensland	1.35	2.65	2.02
Western Australia	2.08	0.37	1.24
Total	1.60	1.38	1.49

a Time expressed in hours.

Table 7.3 Analysis of Variance Results for School Preparation Sessions with Respect to School Type

	F Ratio	Corrected F Ratio	df	Significance Simple	Significance Corrected	R^2
Australian Capital Territory	179.08	44.77	1	$p < 0.001$	$p < 0.001$	26%
Queensland	9.33	2.33	1	$p < 0.05$	ns	2%
Western Australia	3.72	0.93	1	ns	ns	1%

State	Government Schools	Non-Government Schools	Total
Australian Capital Territory	1.31	3.72	1.88
Queensland	1.86	2.45	2.14
Western Australia	0.24	0.45	0.32

School Type Differences in Sessions Attended at School

Table 7.3 reports the results of analyses of variance for school preparation sessions with respect to school type. The analyses were carried out for each State separately. The only significant difference recorded was in the ACT where the students from non-Government schools attended substantially more sessions at school than students who attended Government schools. Although the trend for non-Government school students to attend more preparation sessions at school was evident in all three authorities the differences in Queensland and Western Australia were not significant. It can be seen from the R^2 values that very little of the variance was explained in Queensland and Western Australia.

School Type Differences in Home Preparation Hours

The analysis of variance results for home preparation time in hours with respect to school type are reported in Table 7.4. After design effects were taken into account no significant differences were found between school type in the number of hours preparation students did at home. In both the ACT and Western Australia the model explained less than one per cent of the variance and in Queensland it explained only one per cent. The lack of differences in preparation at home is most likely the consequence of a lack of materials available for such preparation, rather than a recognition of the possible lack of advantage to be obtained from additional preparation under the conditions that some preparation was undertaken at school.

Table 7.4 Analysis of Variance Results for Home Preparation Time[a] with Respect to School Type

	F Ratio	Corrected F Ratio	df	Significance Simple	Significance Corrected	R^2
Australian Capital Territory	0.26	0.07	1	ns	ns	<1%
Queensland	5.19	1.30	1	$p < 0.05$	ns	1%
Western Australia	0.03	0.01	1	ns	ns	<1%

State	Government Schools	Non-Government Schools	Total
Australian Capital Territory	1.25	1.00	1.19
Queensland	1.10	3.10	2.02
Western Australia	1.30	1.14	1.24

a Time expressed in hours.

Overview

In the analyses reported above no significant differences were found between the sexes, State or school type in preparation at home. In the case of school preparation sessions attended by students, significant differences were found between the sexes, females attended more than males. Likewise, some differences were found between States. Queensland students attended on average, just over two sessions; ACT students attended on average, just under two sessions; and Western Australian students attended on average less than one session each. Significant differences were also found between school type in the ACT. A discussion of the reasons for the differences across States and between the sexes is not possible from the data recorded. Whether these relatively small differences between State systems, between school type and between the sexes in preparation practices might have contributed to the observed differences between the sexes in performance on ASAT has not been examined in this investigation. This is not to deny the possibility of such differences having an effect, but rather that the differences in preparation practices did not appear to be of sufficient magnitude to warrant a thorough examination of their effects. However, it would seem important that a more extensive investigation should be carried out to examine whether preparation practices would contribute to differences in attitudes and in turn to differences in performance on ASAT.

CHAPTER 8

ITEM BIAS IN ASAT-L

ASAT is used to moderate between subjects for the purpose of university entrance. This purpose is set out as follows in the test specifications:

> ... its major use is that of scaling school assessment (with respect to scholastic aptitude) both between subjects and between schools. Consequently, its validity must be considered at all times in terms of its strength and robustness in carrying out this task. (ASAT specifications)

The construct of 'scholastic aptitude' is defined as follows:

> The term is taken to mean a capacity to undertake the intellectual pursuits common to scholarship at senior secondary and post-secondary levels. (ASAT specifications)

Any exploration of bias in ASAT should, therefore, keep in mind what the test aims to measure and the manner in which the measures are designed to be used. If ASAT is to be unbiased it must measure the same construct with equal accuracy for both males and females. If the test made members of one group appear less able than their actual capacity, then the test would be biased against that group. Such a bias would make the test invalid and would be detrimental to the group against which it was biased.

There are two different types of approaches for exploring bias in a test. The first uses external criteria and the second, internal criteria. The study of bias in the presence of an external criterion has been called 'bias in selection', and the study of bias in the absence of an external criterion to been called 'item bias' (Ironson and Subkoviak, 1979). The study of sex bias in selection for a test such as ASAT would require a predictive validity study that would determine whether or not ASAT made equally accurate predictions for both male and female students. A study by Müller (1982) has examined the predictive validity of Tertiary Entrance scores calculated by rescaling and statistical moderation using ASAT for performance at the Australian National University. Sex differences in prediction were reported only for the subjects of statistics, computing and geography. No information was available on the predictive power of the ASAT test itself. This procedure was unavailable for the present study because of the lack of external criteria against which to assess ASAT. The present report could only consider item bias.

Methods for detecting item bias have proliferated over the last few years. As discussed in Chapter 2, a range of statistical techniques has been developed to attempt to identify bias using only internal criteria. None of the methods that have been developed so far are completely adequate. Each of the methods adds to the information available about an item and nature of any bias it might contain. This investigation was concerned only with a study of ASAT-L in the Australian Capital Territory, since time imposed severe restrictions on both the number of tests that could be examined and the number of situations in which the investigation could be conducted.

The test was first analyzed using traditional methods. This included the calculation of item facilities (proportion of students getting an item correct), discrimination (point-biserial correlations between the item and the sub-scales to which it was related) and inter-correlations between the sub-scales of the test. The items in the test were then factor analyzed to determine if the factor structure differed for males and females. Although factor analysis has not been recommended as the sole procedure for the detection of item bias (Osterlind, 1982) it could provide useful information and was therefore undertaken in the study. Furthermore, only factor analyses would show if the test score variance was composed of the same theoretical constructs for each sex.

The most advanced techniques for the detection of item bias are the chi-square (Marascuilo and Slaughter, 1981) and the latent trait methods (Ironson, 1983). The chi-square procedure is easy to employ but was not used here. The chi-square procedure divides the sample into strata by ability, in this case aptitude. The proportion of students succeeding on an item within each stratum is compared between the sexes. However, if there were a sufficient number of biased items in the test the total test score might be biased, so that the allocation of students to a stratum might also be biased against a subgroup. In the case of ASAT the females' mean score was significantly less than that for males. If this were a result of test bias then females would be allocated to strata below that for males of equal aptitude, and they would not be expected to perform as well as their true capacity would indicate. Any item bias against females might therefore be disguised by this technique. In preference to the chi-square procedure a latent trait method was employed.

The latent trait methods were considered to be the most strongly based in theory and the most elegant methods for teasing out bias. They had the

important feature that the results were independent of the sample used for calibration. This means that if the distribution of aptitude was different for males and females then the results from applying a latent trait method would be unaffected. Although the three-parameter model might have been preferred the Rasch one-parameter model was employed.

The Rasch model required no assumptions about the comparability of aptitude distributions of different groups. Its conclusions would not be dependent on the characteristics of the sample of persons taking the test since the estimation of item characteristics would be sample invariant. However, the model also made strong assumptions about unidimensionality and equal item discrimination. Any lack of fit to the model would indicate that these assumptions had been violated.

Two methods can be used with the Rasch model to determine if items are behaving differently for the subgroups. The first employs item plots (Osterlind and Martois, 1980), where the difficulty of the item for one group is plotted against the difficulty of the item for the other group. This enables the item invariance to be investigated. It is a measure of the quality of the individual item to be free from, or contaminated by, some degree of bias. The second approach is a comparison of the between group fit for each of the sexes. A high between group fit indicates that the item is not discriminating in the same manner as the rest of the test. It may also indicate that different traits are being measured by the items (Yeh et al, 1980).

Results

Traditional Statistics

The item facilities and point-biserial discrimination indices for males and females are recorded in Appendix 8.1. The majority of Humanities items were easier for females than males. The Social Science items tended to be easier for males while most of the Mathematics and Science items were also easier for the males.

The point-biserial correlations between the items and the full test did not appear to be consistently higher for males or females. Similarly the point-biserial correlations between the items and their subscales did not indicate that items might have discriminated more highly for males or females. The intercorrelations between each of the subscales were calculated for the males and females. The results are reported in Table 8.1. All of the correlations are similar for males and females, indicating a similar relationship between subtests for the sexes.

Table 8.1 Intercorrelations between the Subscales

N = 2432	M	S	SS	H	V	Q	FT
		.685	.522	.430	.490	.887	.777
Mathematics(M)	1.000	.699	.528	.446	.503	.889	.791
		.704	.530	.422	.485	.892	.788
			.645	.593	.681	.912	.891
Science(S)		1.000	.628	.592	.672	.922	.895
			.639	.570	.659	.921	.893
				.602	.762	.707	.813
Social Science(SS)			1.000	.604	.772	.690	.805
				.596	.762	.698	.811
					.951	.550	.822
Humanities(H)				1.000	.947	.557	.817
					.949	.529	.866
						.623	.888
Verbal(V)					1.000	.624	.881
						.604	.875
							.913
Quantitative(Q)						1.000	.919
							.914
Full Test(FT)							1.000

Factor Structure

Two different factor analyses were carried out. Initially a principal components factor analysis of the items was undertaken for each sex. In both cases two factors were identified and rotated using a varimax rotation. Appendix 8.2 reports factor loadings greater than 0.15 for each of the items. For both males and females the first factor was the quantitative subscale and the second factor was the verbal subscale. Most Humanities items loaded clearly on the second factor while the Mathematics items generally loaded on the first factor. The Social Science and Science items did not always load as cleanly on one factor and this was reflected by their allocation to both the quantitative and verbal subscales. In general, items loaded in a similar way for both males and females.

The second factor analysis to be carried out was a principal components analysis of the units which made up the ASAT test. Two factors were rotated using a varimax rotation. These two factors were identified as the

quantitative and verbal subscales. The results of this analysis are reported in Appendix 8.3. The manner in which units loaded was similar for males and females. The Mathematics units loaded on the second factor and the Humanities loaded on the first factor. Some Social Science and Science units loaded on both factors, but this was to be expected since they contained some items classified on the verbal and others on the quantitative subscales.

Rasch Analyses

Using the program BICAL (Wright and Mead, 1976) the full test was calibrated separately for the males and females. Each of the Mathematics, Humanities, Social Science, and Science subscales was also calibrated separately for males and females. From each calibration the difficulty of each item and between group fit were recorded. These results are reported in Appendix 8.4.

To assist with the comparison of Rasch difficulty calibrations, item plots were undertaken. The difficulty of an item for males was recorded on one axis and the difficulty of the item for females was recorded on the other axis. When plotted the distance that an item deviated from the 45° diagonal line indicated the degree to which that item was sample invariant (Wright and Stone, 1980). From an examination of these plots it was found that the Humanities and Social Science items grouped more closely about the 45° line than did the Mathematics and Science items.

The between group fit values are reported in Appendix 8.4. If an item had a high between group fit it would be discriminating in a different manner from the rest of the test. If the between group fit for an item were significantly different for each sex then that item would not be related to the other items in the test in the same way. Two items could have similar point-biserial discriminations but different between group fits because the rest of the items were discriminating differently. However, items with high between group fit would not necessarily be poor discriminators. If an item had high discrimination in relation to the other items it would have a substantial between group fit even though it might be considered a good item. There were some large differences between the results for males and females with respect to between group fit.

Discussion

The results as pointed out do not appear to provide consistent evidence of bias in the test items. Although the female students found the Mathematics and Science sections more difficult the subsequent analyses indicated that if the effects of different ability distributions within the samples were

Table 8.2 Summary of Possible Elements of Bias

Item	Item Facility (%)	r_{pbi} with subtest	Rasch Difficulty (logits)	Between group fit	Factor loadings
		Humanities Units			
1					
2			M		
3					
4		M		M	
5					
6				F	
7			M		
8					
9					
20					
21					
22					
23					
24			M		M
25			F		
26					
27					
28					
29			F		
30			M	M	
31					
81				M	M
82					
83					
84					
85					
86			M		
87					
88					
89	F		F		
		Social Science Units			
43		M			M
44*		M	F		M
45		M			
46	M		M		
47			F		
48					
49		M			M
50	M		M		
59			M		
60				M	
61					
62		M		M	
63			F		
64		F			
65			M		
66		M			
67				F	
68					
69		F			
70			F		

(continued Page 86)

removed, then the items were performing in a similar way for males and females in relation to each other. Table 8.2 is a summary of the results presented in Appendices 8.1 to 8.4. From each of the four subscales the items that behaved significantly differently for males and females were selected. Items marked M could be said to have potential bias in favour of males while items marked F behaved in favour of females.

The selection of items that behaved significantly differently for males and females is somewhat arbitrary. It may be considered that since the responses of all students who sat ASAT in the ACT have been analysed, we would be dealing with a population and consequently significance levels would be meaningless. However, the selection of items that behaved differently had to be based on some criterion. Consequently the following working rules were employed for the purposes of screening the items in order to identify items that showed possible bias.

After taking into account design effects the samples of male and female students were considered to be equivalent to simple random samples of about 200. Using a t-test, items which showed a difference in facility that was significant at the 5 per cent level were selected. For the point-biserial correlation between the item and the subscale to which that item belonged, the design effect was estimated to be about 1.5. Consequently items with point-biserial correlations that differed between the sexes by more than 0.09 were selected as being candidates for items that showed possible sex bias.

For the Rasch difficulty estimates the standard errors for each item were about 0.06 logits. A study by Farish (1984) on the stability of Rasch calibrations has shown that the design effect for the Rasch difficulty is about 1.5. On this basis items that differed in difficulty by more than 0.18 logits were selected. The between group fit values are t-values. For large samples the t distribution approaches the normal distribution. Therefore we may consider the difference between two t distributions as similar to the difference between two normal distributions and this may be approximated by a normal distribution. After considering design effects items with t-values that differed by greater than 2.5 were selected as items that showed possible sex bias.

Items that did not load on a factor as expected are also indicated in the summary chart presented in Table 8.2.

Table 8.2 (continued)

Item Number	Item facility (%)	r_{pbi} with subtest	Rasch difficulty (logits)	Between group fit	Factor loadings
		Science Units			
10*	M		M	M	
11	M				
12		M			
13	M		M		
14	M		M		
15		M		F	
32		M	F		
33	M		M		
34		M	M		
35*	M		M	F	
36			M		
37					
51			F		
52			F		
53					
54			F	M	
55					
75		M	F		
76		F			
77*		M	F	M	
78					
79		F			
80					
94			F		
95			F		
96			F		
97			F		
98			F	M	
99			F		
100			F		
		Mathematics Units			
16			F		
17			F		
18					
19			F		
38	M				
39					
40					
41					
42	M	M			
56					
57					
58					
71	M				
72	M				
73	M				
74	M		M		
90	M		M		
91			F		
92	M		M		
93	M	F			

* Items with bias recorded on three of the five indices.

On most occasions items that had the largest differences in facility also had large differences in Rasch difficulty. Items with different discriminations also generally had different between group fits. If an item was found to favour one group on facility grounds and on Rasch difficulty grounds then it was found to favour the same group. This was not the case for between group fit. For example, Item 15 discriminated more highly for males, yet it had a more satisfactory between group fit for females. This could occur because for the males the item discriminated in a manner different from the rest of the Science items. For the females the item had a discrimination pattern similar to the rest of the Science items.

Four items were selected and discussed regarding the possible reasons for any sex bias within the items. These items were selected on the basis that they were among the items which were seen to have potential for bias on at least three of the five criteria discussed above.

Discussion of Selected Items

Four items on the test had item characteristics which pointed to a possibility of bias in favour of males. They were 44 (Social Science), and 10, 35, 77 (Science). These items are marked in Table 8.2 with an asterisk.

Item 44 is from a Social Science unit that required students to read and interpret two graphs of a type that they were unlikely to have seen before. Item 44 was found to correlate more highly with the verbal subscale while the remaining items in the unit, with the exception of Item 47, correlated more highly with the quantitative subscale. Upon inspection of the items a difference in the questions can be observed. In order to answer the items in the unit it is necessary to look for the solution in the graphical representation. However, Items 44 and 47 involved a use of language that required more than just graph reading skills.

In Item 44 the children leave school as soon as they are permitted. While the graph indicates that the children go straight from school to home, it does not indicate when the students are permitted to leave. The only correct answer in Option D 'is irrelevant to the information in Figure 1 and cannot be supported or refuted by such information'.

This item was found to discriminate more highly for males than females and it was found easier by females than males. It did not load on the verbal factor for females as it did for males.

Item 10 comes from a Science unit that discusses car braking with the use of a graph. Essentially it is a unit that relies on the students' ability to read and interpret graphical material. Four of the six items in the unit were found significantly more difficult by females than males. Three of these items had significantly greater Rasch difficulty estimates for females than males. Item 10 was not only found to be more difficult for females but it also had a less satisfactory between group fit.

Item 35 is a unit that discusses 'basal metabolic rate'. A graph is used in the unit and items require some calculation as well as graph reading. Item 35 was found to be significantly easier for males than females, and this was also reflected in the significantly lower Rasch difficulty estimate for males. The between group fit for the item is significantly better for females than males. Although the whole unit is easier for males than females only Item 35 shows the possibility of bias on three criteria. The only hints from the item content that might indicate a source of bias are the words "female whose body surface". This would appear unlikely to be the source of bias because other items in the same unit also refer to bodies, females, girls and men.

Item 77 comes from a science unit on microbial ecology. It draws on a good deal of stimulus material in written, tabular and graphical form. There appears to be nothing in the item that distinguishes it from the other items in the unit. It was, however, found to discriminate more highly for males, have a lower Rasch difficulty for females and a more satisfactory between group fit for males.

There was nothing about the content or the wording of any of these items which distinguished them as a group, or as a discrete item, with specifically "male" or "female" orientation. To suggest that the mention of cars (Item 10) or a female body (Item 35) indicated a "male orientation" would appear contrary to the fact that other items in these units were not significantly biased. Inferences based on the vocabulary of items would not appear to stand without rather more evidence than provided here.

Summary and Conclusions

Five different indices were used to detect possible bias in items on ASAT-L. Each of the indices identified a number of items that might contain elements of bias. The results from studying these indices could, at best, be called inconclusive. It would appear that the techniques that are currently available for the detection of item bias are inadequate. More research needs

to be undertaken in this field before it can be particularly useful in practical situations. The five indices did not appear to show any consistent sources of bias in difficulty or discrimination. At the content level there was no indication as to why some items were shown to contain elements of bias yet other items based on the same material and of a similar wording did not exhibit bias. From the results presented in this chapter there would appear to be no conclusive argument for sex bias in the items of the version of ASAT used in 1983.

CHAPTER 9

DIFFERENT FACTORS AND THEIR EFFECTS

ASAT is designed to be equally suitable for all Year 12 students regardless of the subjects they are studying. The general structure of the test is prescribed in the test specifications:

> In each edition, questions are grouped in units, based on stimulus material in a variety of forms. Any specific information required to answer the questions is contained in the stimulus material, pitched at a level of knowledge assumed to be common to all students at Year 12. (ASAT Test Specifications)

ASAT currently consists of material drawn in the following proportions; Humanities 30 per cent, Social Science 20 per cent, Mathematics 20 per cent and Science 30 per cent. Care is taken in ensuring that the material presented does not coincide with the specific content of Year 11 or Year 12 syllabuses or courses of study. The basic level of knowledge that is required in the items is generally regarded as equivalent to the Year 10 level. On the basis of this, the role of ASAT is such that performance on the test should not be influenced significantly by variations in high school coursework. If, however, ASAT were found to be sensitive to variations in coursework, after due allowance had been made for student ability and aptitude, then the observed sex differences might be due, in part, to differences in the selection of courses made by male and female students.

Previous research by Pallas and Alexander (1983) has shown that performance on tests such as ASAT could be responsive to differences in coursework. Pallas and Alexander found that 60 per cent of the difference between male and female scores on the SAT-M were a result of boys studying a greater amount of mathematics than girls. Pallas and Alexander controlled for initial mathematical aptitude using the SCAT-Q and then argued that greater exposure to mathematics courses had increased the boys' performance on the SAT-M. Since it has been traditional for males to study a greater amount of mathematics than females (Moss, 1982) the Pallas and Alexander results might have some bearing on the observed differences in performance between the sexes being investigated in this study.

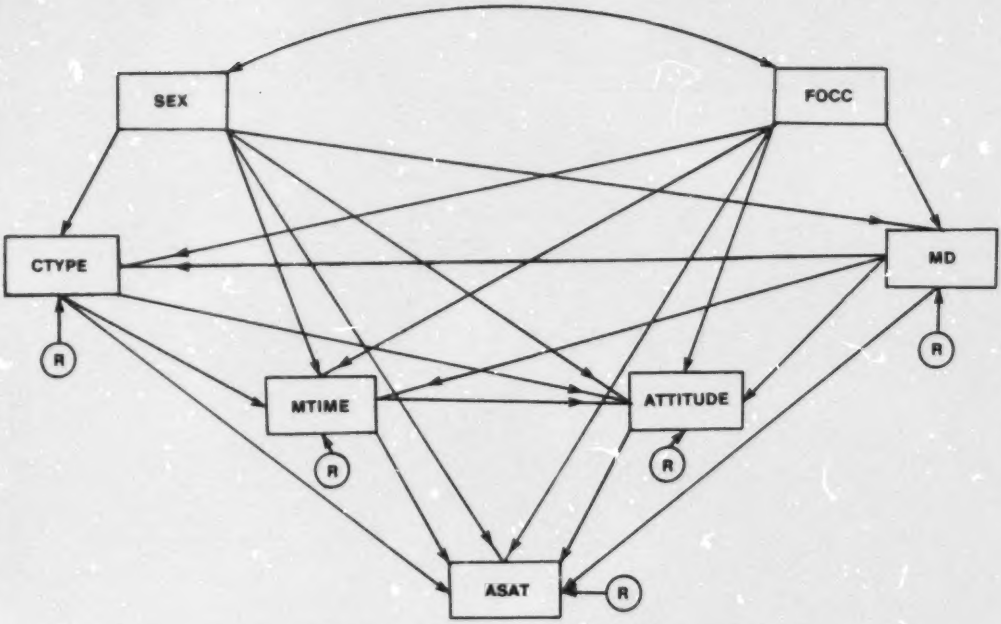

Figure 9.1 A Proposed Causal Model

A First Model

After considering the factors that may have an affect on ASAT scores, a causal model was proposed. This model could be used to explore the relationships between; (1) ASAT performance, (2) Course type, (3) Socio-economic status, (4) Attitudes towards ASAT and (5) Sex. The relationships could be explored using the following variables.

Sex	The sex of each student was available and coded 0, male and 1, female.
Socio-economic status	Father's occupation was used as a measure of SES. A six point scale was used with one high status to six low status. (FOCC)
Course type	For the purpose of these analyses the variable was coded 0 if a student did not study mathematics and 1 if they did. (CTYPE)
Hours mathematics	The number of hours per week that a student spent in the study of mathematics was recorded. Values ranged from 0 to 9. (MTIME)
Attitudes towards ASAT	Four attitudes were used as described in Chapter 6.

Table 9.1 Results of Regression Analyses with Confidence in Success Included.

	Sex	Fathers Occupation	Variable 1	2	3	4	R^2
1. ASAT as a male domain	1.89 (0.23)	-0.02 (-0.01)					6%
2. Course type	-0.03 (-0.06)	0.01 (0.06)	0.00 (0.05)				1%
3. Time studying mathematics	-0.45 (-0.11)	-0.02 (-0.01)	0.02 (0.03)	2.89 (0.40)			17%
4. Confidence in success	-2.00 (-0.20)	-0.47 (-0.13)	0.36 (0.17)	-0.52 (-0.03)	0.32 (0.10)		13%
5. ASAT score	2.82 (0.09)	-0.34 (-0.03)	0.15 (0.04)	-6.69 (-0.10)	2.06 (0.25)	1.80 (0.54)	39%

Table 9.2 Results of Regression Analyses with Effectance Motivation Included.

	Sex	Fathers Occupation	Variable 1	2	3	4	R^2
1. ASAT as a male domain	1.89 (0.23)	-0.02 (-0.01)					6%
2. Course type	-0.03 (-0.06)	0.01 (0.06)	0.00 (0.05)				1%
3. Time studying mathematics	-0.45 (-0.11)	-0.02 (-0.01)	0.02 (0.03)	2.89 (0.40)			17%
4. Effectance Motivation	-0.46 (-0.04)	0.10 (0.02)	0.13 (0.09)	0.06 (0.00)	0.49 (0.18)		4%
5. ASAT score	-0.48 (-0.01)	-1.25 (-0.11)	0.71 (0.17)	-7.66 (-0.12)	2.32 (0.29)	0.63 (0.21)	18%

Table 9.3 Results of Regression Analyses with Attitude Towards Success Included.

	Sex	Fathers Occupation	Variable 1	2	3	4	R^2
1. ASAT as a male domain	1.89 (0.23)	-0.02 (-0.01)					6%
2. Course type	0.03 (-0.06)	0.01 (0.06)	0.00 (0.05)				1%
3. Time studying mathematics	-0.45 (-0.11)	-0.02 (-0.01)	0.02 (0.03)	2.89 (0.40)			17%
4. Attitude towards success	0.54 (0.08)	-0.08 (-0.03)	0.15 (0.18)	-0.01 (0.00)	0.02 (0.01)		5%
5. ASAT Scale	-1.06 (-0.03)	-1.14 (-0.10)	0.71 (0.17)	-7.61 (-0.12)	2.62 (0.32)	0.53 (0.10)	15%

(i) Confidence in success (CS)
(ii) Effectance motivation (EM)
(iii) ASAT as a Male Domain (MD)
(iv) Attitude towards success (AS)

The model that was proposed for the examination of these relationships is presented in Figure 9.1.

For the purpose of studying these relationships, 509 students from the ACT questionnaire sample (See chapter 6) were matched to their ASAT results. It is important to note that the ASAT performance data were those obtained after raw scores had been standardized using the adjustment procedure employed in 1983 (see Chapter 1). For 1983, the raw scores of students in the ACT on ASAT were standardized to a mean of 65 and standard deviation of 15. The male and female means were then calculated and one third of the difference on mean scores between the sexes was added to all female scores. As a result of this each female student was assigned 0.995 extra score points. The male and adjusted female scores were then pooled and restandardized to a mean of 65 and standard deviation of 15.

The model as proposed in Figure 9.1 was examined in five stages.

Stage	Dependent variable	Independent variables
1.	ASAT performance (ASAT)	Sex
		Socio-economic status (FOCC)
		ASAT as a male domain (MD)
		Course type (CTYPE)
		Hours mathematics (MTIME)
		Attitudes towards ASAT
		(i) Confidence in success (CS)
		(ii) Effectance motivation (EM)
		(iii) Attitude towards success (AS)
2.	Attitudes towards ASAT	Sex
	(i) Confidence in success (CS)	Socio-economic status (FOCC)
	(ii) Effectance motivation (EM)	ASAT as a male domain (MD)
	(iii) Attitude towards success (AS)	Course type (CTYPE)
		Hours mathematics (MTIME)
3.	Hours of mathematics (MTIME)	Sex
		Socio-economic status (FOCC)
		ASAT as a male domain (MD)
		Course type (CTYPE)

4.	Course type (CTYPE)	Sex
		Socio-economic status (FOCC)
		ASAT as a male domain (MD)
5.	ASAT as a male domain (MD)	Sex
		Socio-economic status (FOCC)

The three different attitudes concerning how a student felt about ASAT (ASAT as a male domain excluded) were included separately in successive analyses. The model proposed that there was a causal link from ASAT as a Male Domain to the Amount of mathematics students studied. It also proposed that the Amount of mathematics that students studied might affect their Attitudes towards ASAT.

Regression analyses were performed on the data to examine the models for each of the three attitudes. Table 9.1 records the results of the analyses when the Confidence in Success scale was used. Table 9.2 reports the results for the analyses with Effectance Motivation and Table 9.3 presents the results for the Attitude towards Success scale. In each of the tables the metric coefficients are reported along with the standardized coefficients in parenthesis. All coefficients have been included regardless of levels of statistical significance. In general, only standardized regression coefficients in excess of 0.1 were regarded as worthy of further consideration. This value is consistent with the assumption of a design effect of 1.33 for a regression coefficient (Ross, 1976). A standardized regression coefficient of 0.1 is significant at the 5 per cent level under this assumption.

The strongest relationships were found when using the Confidence in Success scale. Using this variable 39 per cent of the variance of ASAT score was explained at the last stage of the analysis and 13 8 per cent of the variance in Confidence in success was explained. A path diagram recording only significant paths is shown in Figure 9.2.

The diagram shows that Sex, Father's occupation, Course type and ASAT as a Male Domain had no significant direct effect on ASAT scores. The only significant causal paths are from Confidence in success and Hours of mathematics. Students who had a higher confidence in success on ASAT performed better on ASAT, as did students who studied more hours of mathematics. Sex had no direct effect on ASAT but females had a significantly lower confidence in success and studied significantly less mathematics. This would decrease

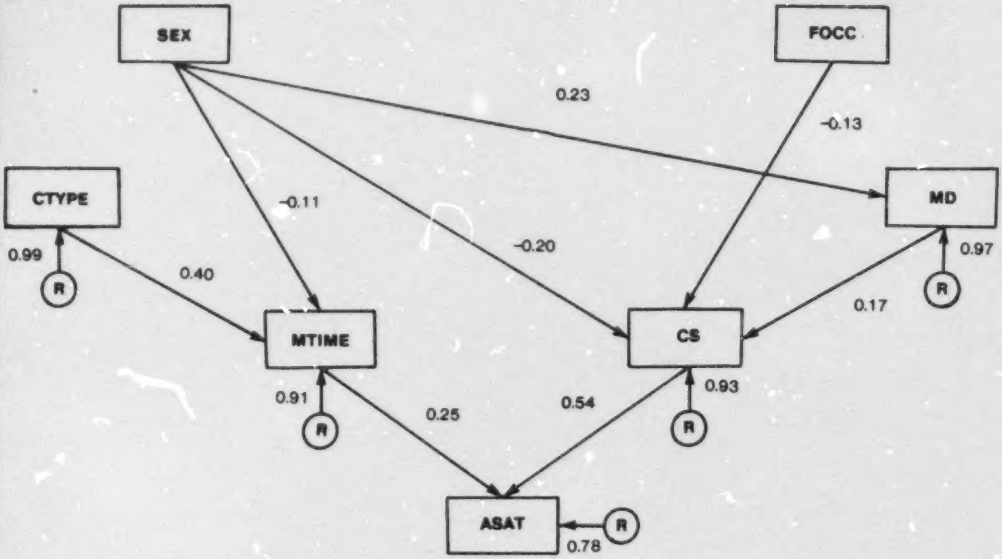

Figure 9.2 Causal Model with Confidence in Success Included[a]

a Standardized coefficients greater than 0.10 are shown.

their ASAT mean in relation to the male ASAT mean. There was some evidence that students from lower socio-economic status families might perform less well on ASAT because of their lower confidence in success.

A Second Model

The above model did not control for the ability of the students. Confidence in success and Hours of mathematics studied might have positive effects on ASAT because it would be the more able students who had higher confidence and who studied more hours of mathematics. The only information that was available for a large number of students that might in some part control for ability was the English assessments. By the inclusion in the model of students' English assessments and Hours of mathematics studied both verbal and quantitative ability could in part be taken into consideration in the regression analysis. The variables Course type and ASAT as a male domain were found to be the least useful in the previous model and were omitted from further analyses. The model proposed is presented in Figure 9.3.

The model as proposed in Figure 9.3 consists of four stages.

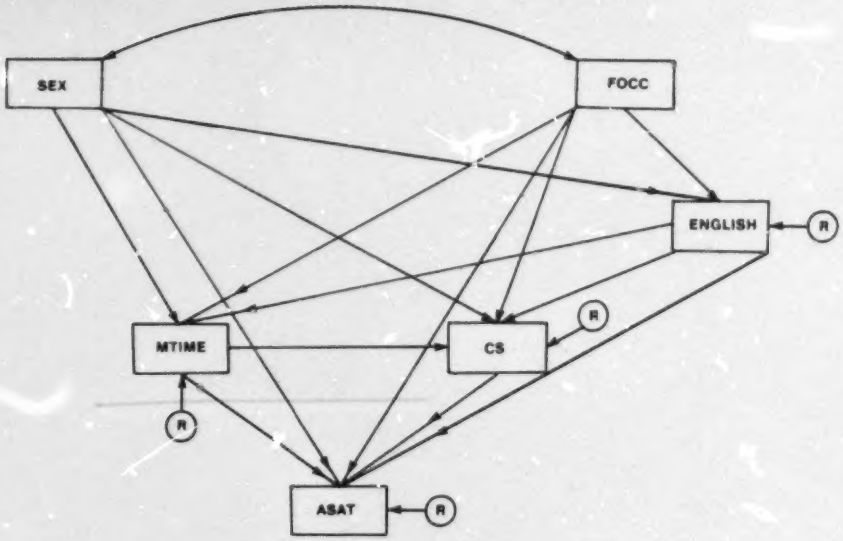

Figure 9.3 A Causal Model Including English

Stage	Dependent variables	Independent variables
1.	ASAT performance (ASAT)	Sex Socio-economic status (FOCC) English (i) Moderated (ENGM) (ii) Unmoderated (ENGC) Maths hours (MTIME) Confidence in success (CS)
2.	Confidence in success (CS)	Sex Socio-economic status (FOCC) English (i) Moderated (ENGM) (ii) Unmoderated (ENGC) Maths hours (MTIME)
3.	Hours of Mathematics (MTIME)	Sex Socio-economic status (FOCC) English (i) Moderated (ENGM) (ii) Unmoderated (ENGC)
4.	English (i) Moderated (ENGM) (ii) Unmoderated (ENGC)	Sex Socio-economic status (FOCC)

Both English assessments direct from the college (ENGC) and English assessments after moderation by ASAT (ENGM) were available for the analyses. Regression analyses for the model were carried out using the variables Moderated English (ENGM) and Unmoderated English (ENGC) successively. The results when Moderated English was used are reported in Table 9.4 and the results when Unmoderated English was used are reported in Table 9.5.

Table 9.4 Results of Regression Analyses with Moderated English

	Sex	Fathers Occupation	Variable 1	Variable 2	Variable 3	R^2
1. Moderated English	4.34 (0.12)	-1.17 (-0.10)				2%
2. Time studying mathematics	-0.57 (-0.14)	0.04 (0.02)	0.02 (0.15)			4%
3. Confidence in success	-2.02 (-0.21)	-0.49 (-0.13)	0.07 (0.23)	0.22 (0.09)		13%
4. ASAT score	0.44 (0.01)	0.17 (0.02)	0.31 (0.34)	1.51 (0.19)	1.37 (0.43)	44%

The results reported in Table 9.4 and Table 9.5 show that the variables Sex and Father's occupation had no significant effects on ASAT performance. The variables which had significant effects on ASAT were, Hours of mathematics, English score and Confidence in success. A comparison of Tables 9.4 and 9.5 shows that the use of moderated English (ENGM) or unmoderated English (ENGC) made no practical difference to the model. This indicates that the moderation procedure using ASAT had little influence on the relative values of college English assessments.

Figure 9.4 shows the significant paths in the causal model when unmoderated English (ENGC) is used as the measure of underlying ability or achievement. The model shows that the sex effects on ASAT are indirect. Female students performed better on English, did less mathematics and had a lower confidence in success. These indirect effects led to the observed differences between the mean scores of males and females.

Table 9.5 Results of Regression Analyses with Unmoderated English

	Sex	Fathers Occupation	Variable 1	Variable 2	Variable 3	R^2
1. Unmoderated English	3.82 (0.14)	-1.05 (-0.11)				3%
2. Time studying mathematics	-0.58 (-0.15)	0.04 (0.03)	0.02 (0.15)			4%
3. Confidence in success	-2.06 (-0.21)	-0.48 (-0.13)	0.09 (0.24)	0.22 (0.09)		13%
4. ASAT score	0.40 (0.01)	0.20 (0.02)	0.37 (0.32)	1.53 (0.19)	1.37 (0.43)	43%

Table 9.6 Means of Variables in the Model

	Male	Female	Difference M-F
Moderated English	62.36	66.19	-3.83
Unmoderated English	66.52	69.82	-3.30
Time studying mathematics	5.85	5.44	0.41
Confidence in success	25.26	23.40	1.86
ASAT score	68.55	66.93	1.62

Accounting for the Observed Difference

For the second model that was proposed complete data were available on 395 students. The means for each of the variables in the model are presented in Table 9.6.

Using the metric coefficients in Table 9.5 at the final stage of the analyses it was possible to calculate the effect of each variable on ASAT score after allowance had been made by regression procedures for all other variables in the model.

The metric coefficient for unmoderated English (ENGC) indicated that for every extra mark a student scored in English they scored 0.37 extra marks on ASAT. It is important to note that this is the case for any student, independent of sex. From Table 9.6 it can be seen that the girls' average unmoderated English mark exceeded the boys' average unmoderated English mark by 3.30 score points. The average ASAT mark of female students therefore, benefited by 1.22 score points. The metric coefficient for Hours spent in learning mathematics indicates that for every extra hour of mathematics studied a student gained 1.52 ASAT score points. Since the average number of hours for boys exceeded the average number of hours for girls by 0.41 hours the males' mean ASAT score was increased by about 0.63 score units. The Confidence in Success attitude scale had a surprisingly large effect on ASAT scores. For every extra point on this scale a students' ASAT score was increased by 1.37 score points. Since the boys' average confidence in success exceeded the girls' average by 1.86 points the resultant advantage on ASAT for boys was 2.55 score points. These three major effects gave the boys' mean an advantage of 1.96 score points over the females for these data. The non-significant coefficient for sex in Table 9.5 of 0.40 gave a total difference in favour of the males of about 1.56 score units.

Without consideration of the effects of a difference between the means of father's occupation this total effect of 1.56 score points in favour of males corresponds to the observed difference on ASAT of 1.62 score points.

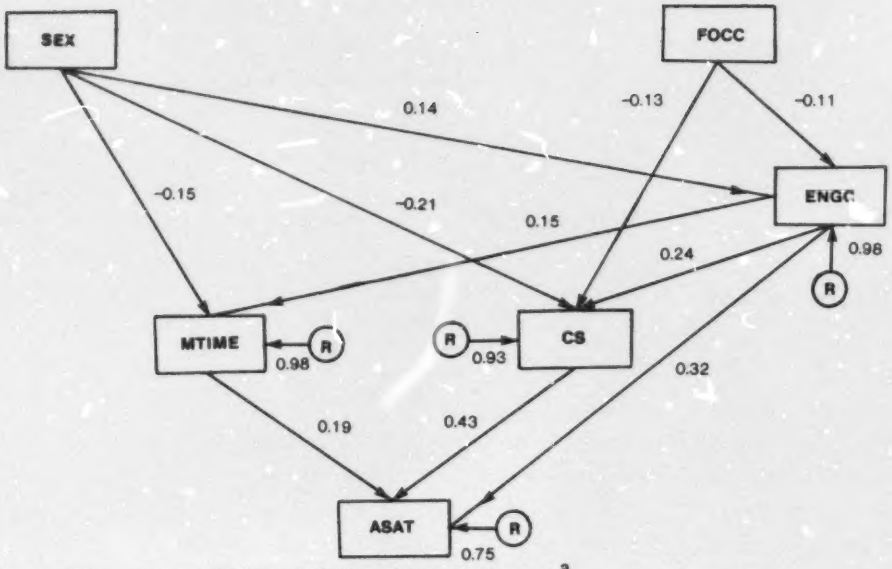

Figure 9.4 Causal Model with Unmoderated English[a]

a Standardized coefficients greater than 0.10 are shown.

Figure 9.5 shows in a diagrammatic form the effects of each of the variables considered in the model. As the standardization procedure used for the 1983 ASAT involved an adjustment that increased the female ASAT mean scores for this sample by an unknown amount it is difficult to determine exactly the effects this had on the results of these analyses. Each female in the ACT was given an extra 0.995 score points before the scores were restandardized to a mean of 65 and a standard deviation of 15. In the sample used for these analyses the difference in favour of females of 0.40 score units is probably explained by the adjustment. If the adjustment had resulted in the girls in the sample having their average score increased by the full amount of the adjustment then the estimated actual sex difference for this sample would be 2.62 score units. As indicated by Figure 9.5 the result would be about 0.60 score units advantage for males rather than the 0.40 advantage in favour of females that was calculated in the analyses.

Summary and Conclusions

In the data in this study, no significant direct effect of sex on performance on ASAT was found. Further sex differences in performance were found to be the result of females having a higher verbal ability as measured by English assessments, males studying more mathematics and males having greater confidence in success on ASAT. For this sample it was found that the female students achieved about 0.4 score points on ASAT more than males, after regressing out socio-economic status, English assessment, hours spent

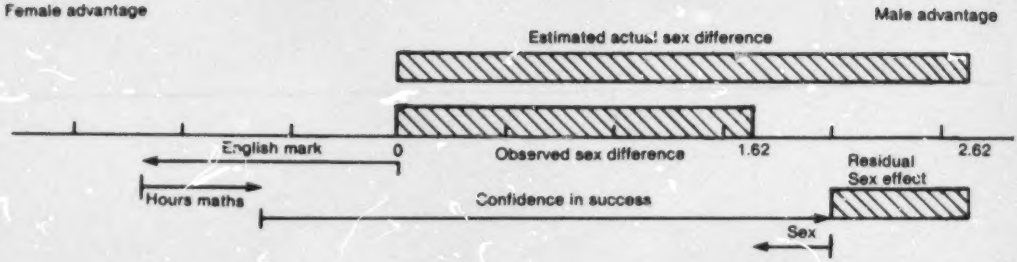

Figure 9.5 Contribution of Variables to the Observed Sex Difference

studying mathematics and confidence in success. The adjustment to ASAT scores in 1983 would certainly account for the positive effect of sex; i.e. being a female increased a student's ASAT score by about 0.40 marks. After allowing for this adjustment of approximately 1.0 score points, being a male increased a student's ASAT score by about 0.60 score points.

ASAT was found to favour students who scored well in English and those students who studied more hours of mathematics. English, in part, tests verbal aptitude and consequently it might be expected that there would be a significant positive effect related to English assessments. The number of hours of mathematics studied is not itself a direct measure of aptitude and consequently it may be considered that the number of hours of mathematics studied provides an advantage for performance on ASAT. However, it is not unreasonable to expect that it is those students of greater quantitative aptitude who would study more mathematics. The significant mathematics hours effect may then be a result of this, since there is no control for quantitative aptitude within the model. It was, however, the very large effect which the score on the Confidence in Success attitude scale had on ASAT performance that was the most surprising. For some undetermined, perhaps social, reason, females had a significantly lower confidence than males.

The relationship between confidence in success and performance might be expected since, in part, the Confidence in Success variable measured how well students thought they went on ASAT. It is still clear however, that after confidence in success is taken into account, sex has no significant effect on ASAT score. The predictors of success are English ability, hours spent in the study of mathematics and confidence.

CHAPTER 10

THE FINDINGS AND THEIR IMPLICATIONS

The purpose of this enquiry was to examine the apparent sex bias in ASAT. This was done by investigating the cause of the bias and providing information that would enable recommendations, where necessary, to be made regarding the procedures employed in the construction, format and use of the test. There was also a need to consider the evidence available from the study for the development of policy to ensure equity between the sexes in the future development of new forms of the test.

Wherever possible within the study, replication was undertaken across the three examining regions that currently use ASAT; the Australian Capital Territory, Queensland and Western Australia. This was done because it was argued that if the findings could be replicated in the three systems, then a high degree of generality might be claimed for such findings. With replication there was the possibility that findings could have the following characteristics:

1 the same results may have been observed in each of the examining regions,
2 results may have been unique to one of the examining regions; and
3 common sense or popular myth might have suggested a result, yet a detailed investigation of the evidence gave no support of such findings.

The findings of this investigation included some results of each of these types.

It was considered necessary to explore the problem by concentrating, concurrently, on five key issues:

1 retention
2 attitudes
3 preparation
4 item bias
5 differential coursework.

As a result of the diversity of these issues, the study had a multifaceted nature, utilizing analyses that ranged from the somewhat arbitrary examination of interview data to the application of statistically sophisticated techniques and mathematical modelling. Despite the range of techniques employed and analyses undertaken, every technique complemented each of the others in the development of a body of evidence regarding the above five issues.

1 The retention issue was studied through regression analyses and the development of a mathematical model to determine the effects of different retention patterns for male and female students on the difference in their ASAT scores.

2. Students' attitudes towards ASAT were explored through four attitude scales that were developed for the investigation, and placed at the end of the questionnaire. Factor analysis was employed to establish the four scales, as was analysis of variance to examine differences between student attitudes. Causal modelling was used to examine relationships between attitudes and performance on ASAT.

3 The issue of preparation was dealt with separately by the interviews and the questionnaire. The interviews focussed on what students believed about the preparation they had received; Was there enough? Was it useful? Would more have helped? Were there adequate preparation materials available? The emphasis of the questionnaire was on the amount of preparation rather than quality or type.

4 The problem of item bias was tackled using both classical test theory and latent trait (item response) theory. ASAT-L, administered in 1983, was analysed to determine if any items exhibited different response patterns for males and females.

5 The differential coursework hypothesis was examined with the interviews, the questionnaire data and ASAT data obtained from the examining authorities. In the interviews students were given the opportunity to express their views on whether they considered that ASAT gave an unfair advantage to students of any particular subject background. Using ASAT score data the relationships between course type and performance were examined with causal modelling.

All of the analyses were seen as necessary and each provided insights into the nature of the apparent sex bias in ASAT.

Results of the Study

A list of the 12 major propositions which have been developed from the study is now presented as a summary of the findings. Each propositional statement is followed by a brief commentary. However, more detailed descriptions and discussions of these results have been provided in the preceding chapters and in the appendices to this report.

Proposition 1. If retention rates for male and female students were equal no significant sex difference could be expected.

Regression techniques were used to examine the relationship between differences in ASAT scores and differences in retention rates. A strong relationship was found to exist between the differences in male and female retention rates and the difference between male and female mean ASAT scores.

On the assumption that an increase in retention rate was a result of more less able students staying on to the Year 12 level, the greater retention rate of females would result in a lower mean ASAT score for females than males. It is expected that this has led to at least part of the observed difference between male and female performance on ASAT.

From regression analyses of the models developed under different assumptions, it was found that it might be expected that males and females would not have significantly different ASAT mean scores if retention rates were equal.

Proposition 2. *The greater retention rate of females explained about half of the observed differences in mean scores of male and female students.*

Using a mathematical model to predict an expected difference in ASAT score, based upon the differences in male and female retention rates, it was found that about half of the observed difference could be explained. That is, the observed differences in ASAT mean scores for males and females were about two times larger than would be expected based on retention differences alone. This relationship is assumed to be a result of the connection between retention rate and the other factors examined in this study such as differences in attitudes and course selection.

Proposition 3. *Students were not aware of any marked sex bias in the test.*

In the interviews that were carried out in the Australian Capital Territory and Queensland, students did not express the view that they believed there was a marked sex bias in ASAT. They could not identify any section of ASAT, based on content or structure grounds, that they thought would lead to an unfair disadvantage for females as a consequence of their sex alone.

It must be kept in mind that these were opinions expressed by students who had recently sat for ASAT and it does not discount the possibility of the existence of bias in the test. It was, however, interesting to note the absence of such opinions among the students regarding bias in the test, especially in the Australian Capital Territory where the tests' validity had recently come into question in the media and through public debate.

Proposition 4. *Female students had significantly less confidence in success on ASAT than male students.*

Analysis of the data provided by the questionnaire regarding students' confidence in success on ASAT indicated that females had significantly less

confidence in their ability to perform well on ASAT, than did the males. This result was supportive of previous research findings that have consistently found that females have significantly less confidence than males in their ability to perform well in areas such as mathematics. Although this result may not be considered surprising it provides evidence that the lack of confidence in success of the female students is a significant factor in relation to the problem of the apparent sex bias.

Proposition 5. Female students felt that they, as a group, were equally suited to performing well on ASAT as males.

An attitude scale was designed to test if students felt that females were suited to performing as well on ASAT as males. Scores on this scale indicated that females felt strongly that they were equipped to perform as well on ASAT as males were. The female scores on the scale were therefore significantly higher than the scores of the males. At first these results may appear contradictory to those that indicated females had a significantly lower confidence in success than the males. However, it should be noted that females' lower confidence in success indicated that they, as individuals, had less confidence in success than the male students, while the higher scores on the ASAT as a Male Domain scale indicated that females felt strongly about their abilities, as a group. In the first of the proposed causal models it was found that high scores on the ASAT as a Male Domain scale led to increased confidence in success and this had positive effects on ASAT performance.

Proposition 6. There were significant differences between the examining regions in the amount of preparation that students had undertaken for ASAT.

In the previous research there is some debate about the effect of preparation on aptitude test scores. As a result of this it was deemed necessary to explore the issue and determine if there were differences between the examining regions in the amount of preparation that was undertaken by students. It was found that significant differences existed between the regions in the number of preparation sessions students attended at school. No information was available regarding the type or quality of that preparation. Since differences were found it is hoped that further research can be undertaken to determine the possible effects of these differences on ASAT performance. The examining authorities need to address the problem of differences in preparation. Is there some way that preparation procedures can be standardized? If coaching is important then the coaching that students achieve should not be left to chance.

Proposition 7. *There were significant differences between the sexes in the amount of preparation that students had undertaken for ASAT.*

A significant difference was found between males and females in the amount of preparation sessions students attended at school. Females did significantly more preparation than males. Such differences were not however detected in home preparation time.

Proposition 8. *It was not possible to identify items that illustrated a consistent pattern of bias against one sex group.*

The responses of all students in the Australian Capital Territory on ASAT-L were analysed with a range of techniques that are available for the detection of item bias. Items were explored for differences in difficulty, differences in discrimination and differences in the contribution of the item to the test score variance. Although each of the methods employed indicated a number of items that contained possible bias there seemed to be no consistent pattern of bias that could be identified as a result of item type or content. It was considered that the techniques currently available for the detection of item bias were of limited use. A detailed study of more than one form of ASAT from a variety of examining regions should be undertaken if the possibility of item bias is to be fully explored.

Proposition 9. *A students' home background as measured by father's occupation had no significant direct effect on ASAT performance.*

From the questionnaire, information was obtained on the father's occupation of students and used as an indicator of socio-economic status (SES). Using a causal model this variable was related to course type, hours of mathematics studied, attitudes towards ASAT, performance in English and performance on ASAT. When the attitude scale Confidence in Success was used it was found that there was no significant direct link between SES and performance on ASAT. However, an indirect effect was found since SES had a significant direct effect on confidence in success which in turn had a direct link with performance on ASAT.

When the attitude scales Effectance Motivation and Attitude towards Success were included in the model, it was found that there was a direct link between father's occupation and performance on ASAT. This would have been a result of the removal of Confidence in Success from the model.

Proposition 10. *Students had a small advantage on ASAT for each extra hour of mathematics they studied.*

For every extra hour of mathematics that students studied, a small increase in ASAT scores was found. It is difficult to determine if this was purely a result of studying extra mathematics or a reflection of the fact that students with a greater quantitative aptitude studied more mathematics.

It was found that females studied fewer hours of mathematics than the males and this explained a small portion of the observed sex difference in ASAT scores.

Proposition 11. *A student's confidence in success had the greatest single effect in increasing performance on ASAT.*

In the final model that was used to explain the components of the observed score differences the students' confidence in success had a surprisingly large effect. Those students who had a higher confidence in success tended to perform more highly on ASAT. Since the male students had a significantly greater confidence in success than the female students, a large amount of the observed difference between male and female mean scores on ASAT can be explained in terms of differences in students' confidence. Since it was the attitudes that had the strongest relationship with ASAT performance, then the difference in attitudes between the sexes should be investigated further.

Proposition 12. *A student's sex had no significant direct effect on ASAT Scores.*

In a model that was proposed to examine the relationship between the various factors investigated and ASAT scores it was found that sex had no significant direct effect. In a model without control for verbal ability it was found that being female had an indirect negative effect through confidence in success and hours of mathematics studied, while it had a positive effect through the ASAT as a Male Domain scale. After college English assessments were included as a control for verbal aptitude, it was found that being female had a negative effect through confidence in success and hours of mathematics, while being female had a positive effect through English.

Although the effects noted above resulted in the observed male mean ASAT score exceeding the female mean ASAT score, sex along does not have a significant influence on ASAT scores. The indicators of superior scores are English ability, confidence in success and experience in mathematics.

Implications of the Study and Its Findings

Implications for Further Research

The study highlighted many areas that should be investigated by further research. Some of the most important issues that arose are discussed in the section which follows.

Previous research had indicated that preparation might have important effects on aptitude test performance. Interviews conducted in the Australian Capital Territory and Queensland supported this and pointed out that there were differences between students in the amount of preparation that they undertook. The questionnaire data as reported in Chapter 7 also illustrated that significant differences existed in student preparation. It would appear important that the relationships between preparation and performance and the effect the reported differences might have on ASAT performance and ASAT's validity should be examined.

In terms of attitudes towards mathematics, previous research has shown that females have what is commonly called a "fear of success". This study attempted to incorporate this notion of "fear of success" into the Attitude towards Success scale, which was one of the four scales developed and included in the questionnaire. However, the attempt was to some degree unsuccessful, as the scale had the lowest reliability and the factor analysis results were less satisfactory for this scale than for any of the other scales. It is therefore recommended that a stronger instrument for "Fear of Success" be developed in an attempt to measure this construct and ascertain if it has any effect on performance on tests such as ASAT.

The issue of retention, within the framework of the study was only investigated in relation to the Australian Capital Territory. It is therefore recommended that this investigation be replicated using data from Queensland and Western Australia in order to establish the strength of the results in other settings besides the ACT.

It is also recommended that the causal model described in Chapter 9 should be replicated for Queensland and Western Australia in order to extend the findings and add to their generalizability.

Finally, and perhaps most importantly, it should be stressed that the study was carried out in the absence of any external criteria. It is therefore recommended that a large scale predictive validity study of ASAT be undertaken. This study would compare the predictive power of ASAT for males and females.

Implications for Theory

The strong relationships observed in this study, between attitudes and ASAT, have major theoretical implications. The findings have added support to a body of evidence that suggests that attitudes may have an extremely powerful impact on performance and more particularly on observed sex differences in performance.

If a causal link between attitudes and performance could be established it would have enormous implications not only for the explanation of sex differences but also for the education process in general.

In the Canberra setting, direct sex differences were found to be slight. This indicated that biological differences might be of relatively little importance in comparison with the more powerful societal and attitudinal factors.

The findings on retention rates also pointed to an important issue. Often, average group performance is studied and comments are made regarding student performance on this information alone. The evidence from the study of the effects of retention rate differences, presented in Chapter 4, draws attention to the fact that there may be problems with this approach. Studying mean ASAT scores alone would have shown that females were not performing as well as males, yet after the consideration of retention rate differences it was found that females might well be performing as well as males. There needs to be developed an accepted model for the study of group achievement relationship that takes into account the selection factors that are operating in Australian upper secondary schools and that influence the size of the student group retained at school beyond the age of compulsory school attendance. Without such a model it is likely that misleading statements will continue to be made on group performance.

Implications for Policy

This study arose out of a need to investigate thoroughly the observed difference between the performance of males and females on ASAT. This difference was labelled a 'sex bias' against females undertaking the test. Evidence from this report has suggested that students are not advantaged or disadvantaged on the basis of their sex, after variables such as confidence in success, hours spent studying mathematics and assessments in English were taken into account. However, females as a group were found to have a lower mean score on ASAT largely because they showed much less confidence in success than did males. This pointed to the fact that in the sample studied

in this investigation, the largest single factor that advantaged or disadvantaged students' performance on ASAT was their amount of confidence in success.

If a group of students studying a particular subject were made up exclusively or largely of females, then their mean ASAT scores might be expected to be lower than for groups made up exclusively or largely of males. The ASAT moderating procedure may disadvantage a school subject group that largely consists of females if the same relationships between confidence in success and performance are not occurring uniformly within the whole group which studies that subject. The study indicated that less confidence in success depressed performance on ASAT. If it can be said that confidence in success also affected achievement on a particular school-assessed subject, then it cannot be said that ASAT is disadvantaging that group. If however, the relationship between attitudes (confidence in success) and aptitude (as measured by ASAT) is not operating similarly between attitudes and achievement in a school subject, then the use of ASAT might have disadvantaged that group.

The results obtained have shown that students studying mathematics may be one group that was advantaged by ASAT. There was a significant relationship between the number of hours students spent studying mathematics and performance on ASAT. This may be considered to be a bias in favour of students studying mathematics. However, the study did not control for quantitative aptitude and it might have been that the relationship recorded was a result of the higher quantitative aptitude of those students who studied more mathematics.

Conclusions

The results of educational research rarely provide answers of unequivocal certainty. There will always be a degree of uncertainty in the results of social research, which involves the investigation of highly complex issues. However, the major function of such research should be to highlight ways in which policy can be improved for the benefit of the educational system as well as for the students on an individual basis.

It was found in this investigation that differences between the sexes in retention rates to the Year 12 level could be considered to account for approximately half the difference between the sexes in observed scores on ASAT. It was also found in subsequent analyses that such factors as difference between the sexes in English ability, time spent in the study of

mathematics and differences in confidence in success on ASAT were statistically significant and accounted for substantial proportions of the sex difference in performance on ASAT. However, it is not possible with the data which were available from the survey of students for one year only to disentangle the relationships between retention and these factors of English ability, study of mathematics and confidence in success. Nevertheless it would seem likely that these factors are related to retention. As a consequence it is apparent that somewhat less than half of the sex difference in performance on ASAT remains unaccounted by these factors of retention, English ability, time spent studying mathematics and confidence in success and that the unexplained part was not found to be statistically significant.

Since the major conclusion of this investigation was that a student's sex had no significant direct effect on performance on ASAT, it would appear questionable to adjust the scores of individual students for bias in ASAT. However, the appropriateness of an adjustment in relation to subject or school groups is unclear. In some groups that have a high proportion of, or that are exclusively composed of, females the mean ASAT scores may be depressed by the lower confidence in success of females in general.

Several questions must be asked. Whether an adjustment because of this lower confidence in success is justified? Furthermore, what of the male students who have a lower confidence in success? Do the single sex boys' schools have an unfair advantage?

Drawing causal inferences from non-experimental investigations involves many problems and it is only when particular findings have been replicated that one can have confidence in their validity. Many of the results reported in this study are tentative until further investigations have been undertaken. However, such further intensive investigation should prove rewarding.

BIBLIOGRAPHY

Adams, R.J. Sex Differences in ASAT and Retentivity. Paper presented at Australian Association for Research in Education conference, Canberra: November 1983.

Aiken, L.R. Research on attitudes towards mathematics. Arithmetic Teacher 1972, 19, 229-234.

Aiken, L.R. Two scales of attitudes towards mathematics. Journal for Research in Mathematics Education, 1976, 5, 67-71.

Aiken, L.R. 'Update on Attitudes and other Affective Variables in Learning Mathematics. Review of Educational Research, 1976, 46(2), 293-311.

Alderman, D.L. and Powers, D.E. The Effects of School Preparation on SAT-Verbal Scores. College Entrance Examination Board. Research and Development Report RDR-78-79 No. 4. Educational Testing Service, Princeton N.J., 1979.

Alexander, K.L. and Pallas, A.M. Reply to Benbow and Stanley. American Educational Research Journal, 1983, 20(4), 475-477.

Anastasi, A. Differential Psychology. New York: Macmillan, 1958.

Angoff, W.H. (Ed). The College Board Examinations Testing Program: A Technical report on research and development activities relating to the scholastic aptitude test and achievement tests. College Entrance Examination Board. Princeton, NJ, 1971.

Angoff, W.H. "A Technique for the Investigation of Cultural Differences." Presented at the Annual Meeting of the American Psychological Association, May 1972. (ED 069686)

Angoff, W.H. and Ford, S.F. "Item-race interaction on a test of scholastic aptitude." Journal of Educational Measurement, 1973, 10, 95-105.

Anttonen, R.G. "A Longitudinal Study in Mathematics Attitudes." Journal of Educational Research, 1969, 62(10), 467-471.

Armstrong, J.M. "Achievement and Participation of Women in Mathematics: Results of two national surveys. Journal for Research in Mathematics Education, 1981, 12, 356-372.

Australian Bureau of Statistics. Estimated Resident Population by Sex and Age: States and Territories of Australia, June 1971 to June 1981.

Australian Council for Educational Research. Commonwealth Secondary Scholarship Examination Research Report No. 3: The effects of Coaching for the Commonwealth Secondary Scholarship Examination, 1974.

Bangert-Drowns, R.L., Kulik, J.A. and Kulik, C.C. Effects of Coaching Programs on Achievement Test Performance. Review of Educational Research, 1983, 53(4), 571-585.

Benbow, C.P. and Stanley, J.C. Sex Differences in Mathematical Ability: Fact or Artifact? Science, 1980, 210, 1262-1264.

Benbow, C.P. and Stanley, J.C. Differential Course Taking Hypothesis Revisited. American Educational Research Journal, 1983, 20(4), 469-573.

Birnbaum, A. Some latent trait models and their use in inferring an examinees ability. In M.L. Lord and M.R. Novick, Statistical Theories of Mental Test Scores, Reading, Mass: Addison-Wesley, 1968.

Bookman, A.B. and Iwanicki, E.F. The Effects of Method of Test Preparation on standardized mathematics test performance. Journal of Research and Development in Education, 1983, 16(4), 46-51.

Bourke, S.F. Do you like Maths? Development of a questionnaire for assessing primary school student attitudes to maths. Paper presented at the Australian Association for Research in Education Conference, Canberra, 1983.

Cordall, C. and Coffman, W.E. A method for comparing the performance of different groups on the item in a test. College Entrance Exam Board. Research and Development Report RDR-64-5 No. 9. Educational Testing Service Research Bulletin RB-64-61, 1964.

Carey, G.L. Sex Differences in Problem Solving. "Performance as a function of Attitude Differences." Journal of Abnormal and Social Psychology 56(2) 256-260, 1958.

Clay, R.W. The effect of Student Gender on Performance in a Final year high school physics examination. Search 14, No. 9-10, 273-274.

Cleary, T.A. and Hilton, T.L. An investigation of item bias. Educational and Psychological Measurement, 1968, 29, 61-75.

Coffman, W.E. Sex Differences in responses to items in an aptitude test. Eighteenth Yearbook National Council on Measurement in Education, 1961, 117-124.

Coffman, W.E. and Parry, M.E. The Effects of an accelerated reading course on SAT-V Scores. Personal and Guidance Journal, 1967, 46, 292-296.

College Entrance Examination Board. Effects of Coaching on Scholastic Aptitude Test Scores. New York College Entrance Examination Board, 1968.

Comber, L.C. and Keeves, J.P. Science Achievement in Nineteen Countries. Stockholm: Almqvist and Wiksell, and New York: Halsted (Wiley).

Cooney, G.H. Standardization procedures involving moderator variables - Some Theoretical Considerations. Australian Journal of Education, 1975, 19(1), 50-63.

Cronbach, L.J. Coefficient alpha and the internal structure of tests. Psychometrika, 1951, 16, 297-334.

Dear, R.E. The Effects of a Program of Intensive Coaching on SAT Scores. Research Bulletin RB-58-5, Educational Testing Service, Princeton, New Jersey, 1958.

Diamond, E.E. The AMEG Commission Report on Sex Bias in Achievement Testing. Measurement and Evaluation in Guidance. 1980, 13(3), 135-147.

Doig, M. Students Attitudes to Writing. Paper presented at AARE Conference Canberra, November 1983.

Donlon, T.F. Content Factors in Sex Differences on Test Questions. Research Memorandum RM-73-28. Princeton, N.J. Educational Testing Service, 1973.

Donlon, T.F., Ekstrom, R.B. and Lockheed, M.F. The Consequences in the Content of Major Achievement Test Batteries. Final Report. 1977 ED151415.

Donlon, T.F., Ekstrom, R.B., and Lockheed, M.E. The Consequences in the Content of Major Achievement Test Batteries. Measurement and Evaluation in Guidance, 1979, 11(4), 202-216.

Dorans, N.J. and Kulick, E. Assessing Unexpected Differential Item Performance of female candidates on SAT and TSWE forms administered in December 1977: An application of the Standardization Approach. Research Report RR.83-9. Educational Testing Service, New Jersey, 1983.

Dyer, H.S. "Does Coaching Help?" College Board Review, 1953, 19, 331-335.

Edwards, A.L. Experimental Design and Psychological Research. Holt Rinehart Winston, New York, 1968

Eells, K., Davis, A, Havighurst, R.J., Herrick and Tyler, R. Intelligence and Cultural Differences. Chicago: University of Chicago Press, 1951.

Farish, S. The Stability of Item Parameters. (In press) 1984.

Fennema, E. Mathematics Learning and the Sexes: A Review. Journal for Research in Mathematics Education. 1974, 5, 126-139.

Fennema, E. Sex Related Differences in Mathematics Achievement: Myths, Realities and related Factors. 1977. (ED 160445)

Fennema, E. and Sherman, J. Fennema-Sherman Mathematics Attitude Scales. Catalogue of Selected Documents in Psychology, 1976, 6(1), 31.

Fennema, E. and Sherman, J. Sex Related Differences in mathematics achievement, spatial visualization and affective factors. American Educational Research Journal, 1977, 14(1), 51-71.

Fennema, E. Women and Mathematics: Does Research Matter. Journal for Research in Mathematics Education, 1981, 12, 380-85.

Ferguson, G.A. Statistical Analysis in Psychology and Education. (5th ed) McGraw-Hill, 1981.

French, J.W. The Coachability of the SAT in Public Schools. Research Bulletin, Educational Testing Service, Princeton, New Jersey. 1955.

French, R.J. and Littell, R.C. SAS for Linear Models: A Guide to the ANOVA and GLM procedures. SAS Institute Inc. 1981.

Glickman, J.R. 1976. Cited in Murphy, R.L. Sex Differences in Examination Performance: Do These Reflect Differences in Ability or sex-role stereotypes. 1976. (ED 154265)

Harmann, H.H. Modern Factor Analysis. 3rd Ed. Chicago. University of Chicago Press, 1976.

Hauser, R.M. Socio-Economic Background and Educational Performance. American Sociological Association, Washington. 1971

Helm, M. and Werner, T. A Review of school certification and tertiary entrance procedures in Australia. Paper presented and the Conference of Directors-General of Education, Darwin, 1983.

Hicks, M.M., Donlon, T.F., and Wallmark, M. Sex Differences in item responses on the Graduate Record Examination. Paper presented at the annual meeting of the National Council on Measurement in Education. San Francisco, 1976.

Hilton, T.L. and Berglund, G.W. Sex Differences in Mathematics Achievement - A Longitudinal Study. Journal of Educational Research, 1974, 67, 231-237.

Horner, M. Achievement Related Conflicts in Women. In M. Mednick and S. Tangri (Eds) A new perspective on Women. Journal of Womens Issues, 1972, 28, 157-175.

Hull, C.H. and Nie, N.H. (Eds). SPSS Update 7-9. New York: McGraw-Hill, 1981.

Husen, T. (Ed). International Study of Achievement in Mathematics. Vols.1 and 2. New York: John Wiley & Sons, 1967.

Ironson, G.H. and Siboviak, M. A Comparison of Several Methods of Assessing Item Bias. Journal of Educational Measurement, 1979, 16(4), 209-226.

Iverson, G.R. and Norpoth, H. Analysis of Variance. Beverley Hills, Sage, 1976.

Jacklin, C.N. "Epilogue" in M. Andrissin Wittig and Anne C. Peterson (Eds) Sex Related Differences in Cognitive Functioning. New York: Academic Press, 1979.

Jackson, R. The Scholastic Aptitude Test: A Response to Slack and Porter's Critical Appraisal. Harvard Educational Review, 1980, 5(3), 382-391.

Jensen, A.R. Bias in Mental Testing. Methuen, London, 1982.

Karmel, P.H. Education in the Eighties. Australian Education Review No. 15, ACER, Melbourne, 1981.

Keeves, J.P. Students' Attitudes Concerning Mathematics. Unpublished M.Ed Thesis, University of Melbourne, 1966.

Keeves, J.P. Educational Environment and Student Achievement. ACER, Melbourne, 1972.

Keeves, J.P. Differences between the sexes in mathematics and science courses. International Review of Education, 1973, 19, 47-64.

Keeves, J.P. and Read, A.D. Sex Differences in Preparing for scientific occupations. IEA (Australia) Report 1974:6. Hawthorn, Vic. ACER. 1974.

Keeves, J.P., McBryde, B. Bennet, L.A. The Validity of Alternative Methods of Scaling School Assessment at the HSC Level for the Colleges and High Schools of the ACT. Paper presented at the Australian Association for Research in Education conference, Canberra, November 1977.

Kim, J.O. and Mueller, C.W. Introduction to Factor Analysis. What it is and how to do it. Beverley Hills, Sage, 1978.

Kim, J.O. and Mueller, C.W. Factor Analysis: Statistical Methods and Practical Issues. Beverley Hills, Sage, 1978.

Kish, L. Survey Sampling. New York: John Wiley. 1975.

Kolen, M.J. and Hoover, H.D. The Reliability of Selected Item Bias Procedures. 1982, ED220474.

Kulik, J.A., Bangert, R.L. and Kulik, C. Cited in Bangert-Drowns, R.L., Kulik, J.A. and Kulik, C. Effects of Coaching Program on Achievement Test Performance. Review of Educational Research, 1983, 53(4), 571-585.

Leder, G.C. 1979. Cited in Moss, J.D. Towards Equality: Progress by Girls in Mathematics in Australian Secondary Schools. Hawthorn, ACER, 1982.

Leder, G. Girls and Mathematics: What of the future? In D. Williams (ed.) Learning and Applying Mathematics. Australian Association of Maths Teachers. 1978.

Linn, R.L., Levin, M.V., Hastings, C.N. and Wardrop, J.L. An Investigation of item bias in a test of reading comprehension. Technical Report No. 163,

Lockhead, M.E., Ekstrom, R.B. and Harris, A.M. Sex Discrimination in Education. A Literature Review and Bibliography. Educational Testing Service Research Bulletin, May 1977. (RB 77/5)

Luchins, E.H. and Luchins, A.S. Letters. Science. 1981,212, 117-118.

McDonnell, W. Testing for Student Selection and Tertiary Level. ACER, Melbourne, 1975.

McGaw, B. Admission of Students to Tertiary Study. Australian Journal of Education, 21(3), 1977, 209-225.

McGaw, B. Combining School-based and external assessments of performance to determine University admissions. Paper presented to the ninth international conference of the International Association for Educational Assessment, Blantyre, Malawi, 1983.

Maccoby, E.E. The Development of Sex Differences. Stanford, California: Stanford University Press, 1966.

Maccoby, E.E. and Jacklin, C.N. The Psychology of Sex Differences. Stanford, California. Stanford University Press, 1974.

Marascuilo, L.A. and Slaughter, R.E. Statistical Procedures for analysing item bias based on χ^2 statistic. Journal of Educational Measurement. 1981, 18, 229-248.

Messick, S. and Jungeblut, A. Time and Method in Coaching for the SAT. Psychological Bulletin, 1981, 89(2), 191-216.

Milton, G.A. "The effects of Sex Role Identification Upon Problem Solving Skill." Journal of Abnormal and Social Psychology, 1957, 55(2), 208-212.

Moss, J.D. Towards Equality: Progress by Girls in Mathematics in Australian Secondary Schools. Melbourne, ACER. 1982.

Müller, W.J. Prediction of Student Achievement at ANU from ACT College Assessments. Master Thesis, Australian National University, 1982.

Neale, D.C. The role of attitudes in learning mathematics. Arithmetic Teacher, 1969, 16, 631-640.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K. and Bent, D.H. Statistical Package for the Social Sciences. (2nd ed) New York: McGraw-Hill, 1975.

Novick, M.R., and Jackson, P.H. Statistical Methods for Educational and Psychological Research. New York: McGraw-Hill, 1974.

Nunnally, J.C. Psychometric Theory. New York: McGraw-Hill, 1978.

Oppenheim, A.N. Questionnaire Design and Attitude Measurement. London: Heinemann, 1966.

Osterlind, S.J. and Martois, J.S. Latent trait theory application to test item bias methodology. 1980. (ED 206650)

Osterlind, S.J. Test Item Bias. Beverley Hills: Sage, 1982.

Pallas, A. and Alexander, K. Sex Differences in Quantitative SAT Performance: New Evidence on the Differential Coursework Hypothesis. American Educational Research Journal, 1983, 20(2), 165-182.

Pike, L.W. Short-term instruction, Test wiseness and the Scholastic Aptitude Test. College Entrance Examination Board, New York, 1979.

Powers, D.E. Estimating the effects of Various Methods of Preparing for the SAT. College Board Report 82-2, Educational Testing Service, 1982.

Reid, N. and Gilmore, A. Test Bias! Test Bias? 1983. Item 16, Set No. 2.

Ross, K.N. Searching for Uncertainty. Occasional Paper No. 9, Melbourne, ACER, 1976.

Rounds, J.B. and Hendel, D.D. Mathematics, Anxiety and Attitudes Towards Mathematics. Measurement and Evaluation in Guidance, 1980, 13(2), 83-89.

Rudner, L.M., Getson, R.G. and Knight, D.L. A Monte Carlo Comparison of Seven Biased Item Detection Techniques. Journal of Educational Measurement, 1980, 17(1), 1-10.

Ryan, J.J. Cited in Neale, D.C. The role of attitudes in learning mathematics. Arithmetic Teacher, 1969, 16, 631-640.

Saaris, T.N., Jacklin, C.N. and Tittle, C.K. Sex Role Stereotyping in the Public Schools. Harvard Educational Review, 1973, 43(3), 386-416.

Scheuneman, J. A Method of Assessing Bias in Test Items. Journal of Educational Measurement, 1979, 16(3), 143-152.

SAS Users Guide: Statistics, SAS Institute , Cary, 1982.

SAS Users Guide: Basics. SAS Institute, Cary, 1982.

Senk, S. and Usiskin, Z. Geometry proof writing: A new view of sex differences in mathematics ability. American Journal of Education. 1983, 91(2), 187-201.

Shelley, N. Women and Mathematics in Learning and Applying Mathematics. (Ed). Doug Williams. Australian Association of Mathematics Teachers, 1978.

Shepard, L., Cannilli, G., and Avenill, M. Comparison of Procedures for detailing Test Item Bias with both Internal and External Ability Criteria. Journal of Educational Statistics, 1981, 6(4), 317-376.

Sinnott, L. Differences in Item Performance Across Groups. Research Report 80-19, Educational Testing Service, New Jersey, 1980.

Sinnot, L. The Identification of Biased Items. Research Report 82-3, Educational Testing Service, New Jersey, 1982.

Slack, W.V. and Porter, D. The Scholastic Aptitude Test: a Critical Appraisal. Harvard Educational Review, 1980, 50(2), 154-175.

Slack, W.V. and Porter, D. Training, Validity and the Issue of Aptitude: A Reply to Jackson. Harvard Educational Review, 1980, 50(3), 392-401.

Smillie, K.W. An Introduction to Regression and Correlation. Academic Press, New York, 1966

Stage, E.K. and Karplus, R. Letters. Science, 1981, 212, 114.

Stein, A.A. and Bailey, M.M. The Socialization of Achievement Orientation in Females. Psychological Bulletin, 1973, 80, 345-366.

Strassberg-Rosenberg, B. and Donlon, F. Content Influences on Sex Differences in Performance on Aptitude Tests. ED110493

Sturman, A. Patterns of School Organization: Resources and Responses in Sixteen Schools. Staffing and Resources Study Report No.3. ACER, Melbourne, 1982.

Swafford, J. Sex Differences in first year algebra. Journal for Research in Mathematics Education. 1980, 11, 335-346.

Swinton, S.S., Wild, C.L. and Wallmark, M.M. Investigation of Practice Effects on item type in the Graduate Record Examination Aptitude Test. RR-82-56 Educational Testing Service, New Jersey, May 1983.

Tittle, C.K. Women and Educational Testing. Phi Delta Kappan, 1983.

Tomizuka, C. and Tobias, S. Letters. Science, 1981, 212, 114.

Whitla, D.K. Effect of Tutoring on Scholastic Aptitude Test Scores. Research and Guidance Journal, 1962, 41(1), 32-37.

Wilson, N. A Study of Test-Retest and of marker reliabilities of the 1966 Commonwealth Secondary Scholarship Examination. Information Bulletin Series No. 50. Hawthorn, Victoria, ACER, 1969.

Wise, L.L., Steel, L. and MacDonald, C. Origins and Career Consequences of Sex Differences in High School Mathematics Achievement. American Institute for Research in the Social Sciences. Palo Alto, California, 1979. (ED120846)

Wolleat, P.L., Pedro, J.D., Becker, A.D. and Fennema, E. Sex differences in high school students causal attribution of performance in mathematics. Journal for Research in Mathematics Education, 1980, 11(5), 347-355.

Wright, B.D. and Mead, R.J. BICAL: Calibrating items with the Rasch model. Research Memorandum No. 23. Statistical Laboratory, Department of Education, University of Chicago, 1976.

Wright, B.D., Mead, R. and Draba, R. Detecting and Correcting Test Item Bias with a logistic response model. Research Memorandum 22. Statistical Laboratory. Department of Education, University of Chicago, 1976.

Wright, B.D. and Stone, M.H. Best Test Design. MESA, Chicago, 1979.

Yeh, J.P. and Conklin, J. Using the Rasch Model to examine item bias. (ED 205577), 1980.

Younger, M. A Handbook for Linear Regression. Duxbury Press. Belmont, 1979.

APPENDIX 4.1

Notes on the Calculation of the Sizes of the Age Cohorts

The data in Table 4.1 are for those students who are retained to Year 12 within the ACT Schools Authority System and sit ASAT to obtain a Tertiary Entrance Score. Mature age and special literacy students have been excluded.

The age cohorts are calculated in the following way:

Example If 1000 males are retained to Year 12, made up of

692 17 year-olds
and 308 18 year-olds

then 69.2% are 17-year-olds
and 30.8% are 18-year-olds

if there are 1950 17-year-olds in the ACT
and 2100 18-year-olds in the ACT

then 69.2% of 1950 = 1349
30.9% of 2100 = 649
1998

So the male age cohort is estimated to be 1998.

The data for the number of students in the age groups for the years 1977 to 1982 were taken from ABS Catalogue 3201.0. Estimated Resident Population by Sex and Age. States and Territories of Australia. June 1977 to June 1982. The 1983 data are estimates obtained directly from the Australian Bureau of Statistics.

The Truncated Normal Distribution

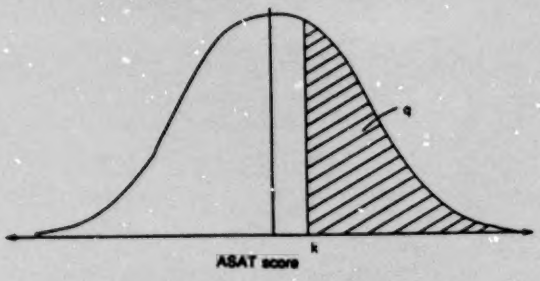

Figure 1. A Hypothetical Distribution of ASAT Scores for the Age Cohort, showing the proportion q above the cut-off point, k.

Suppose that in each year the male age cohort has its ASAT scores following N(0,1) distribution.

Suppose only a proportion, q, of the age cohort has remained to sit ASAT and they form the upper tail as shown in Figure 1.

$$\text{let } \psi(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$$

$$\text{then } q = \int_k^\infty \psi(u)\,du$$

$$\bar{X} = E(X) = \frac{\int_k^\infty u\psi(u)\,du}{\int_k^\infty \psi(u)\,du}$$

$$\text{but } \frac{d\psi(u)}{du} = -\frac{1}{\sqrt{2\pi}} ue^{-u^2/2} \qquad (1)$$

$$\text{so } \frac{1}{\sqrt{2\pi}} \int_k^\infty ue^{-u^2/2}\,du = \frac{1}{\sqrt{2\pi}} \int_\infty^k -ue^{-u^2/2}\,du$$

$$= [\psi(u)]_\infty^k$$

$$= \psi(k)$$

$$\text{So } \bar{X} = \psi(k)/q = y/q$$

$$qE(X^2) = \int_k^\infty u^2 \, \phi(u)du$$

$$= \int_k^\infty u.u \, \phi(u)du$$

From equation (1)

$$= \left|-u \, \phi(u)\right|_k^\infty - \int_k^\infty -\phi(u)du$$

$$= k\phi(k) + q$$

$$= ky + q$$

So $E(X^2) = k^y/_q + 1$

$$var(X) = E(X^2) - E(X)^2$$

$$= k^y/_q + 1 - {}^{y^2}/_{q^2}$$

$$= 1 - {}^y/_q \left({}^y/_q - k\right)$$

Suppose the female age cohort also have their ASAT scores following an N(0, 1) distribution.

If a proportion q_f remain to sit ASAT and they form an upper tail as shown in Figure 1 then

$$E(X_f) = \bar{X}_f = {}^{y_f}/_{q_f} \qquad E(X_f^2) = k_f \, {}^{y_f}/_{q_f} + 1$$

$$var\ (X_f) = 1 - {}^{y_f}/_{q_f} \left({}^{y_f}/_{q_f} - k_f\right)$$

where k_f, y_f and q_f are defined as k, y and q above.

If we pool the male and female population and using obvious notation

$$E(X_{m+f}) = \frac{q_m\bar{X}_m + q_f\bar{X}_f}{q_m + q_f} = \frac{y_m + y_f}{q_m + q_f}$$

$$E(X_{m+f}^2) = \frac{q_m(k_m y_m/q_m + 1) + q_f(k_f y_f/q_f + 1)}{q_m + q_f}$$

$$= 1 + \frac{k_m y_m + k_f y_f}{q_m + q_f}$$

$$\text{var}(X_{m+f}) = E(X^2_{m+f}) - E(X_{m+f})^2$$

Assume that ASAT score, X, and a retentivity variable, R, have a joint bivariate normal distribution.

If R is marginally N(0, 1) and the students that remain to year 12 and sit ASAT form the upper tail of the R variable.

For the conditional probability density function of X given R = r

$$E(X|R=r) = \rho r \qquad E(X^2|R=r) = 1 = \rho^2 + \rho^2 r^2$$

where $\rho = \text{corr}(X, R)$

From previous calculations

$$E(RI_{R\geq k}) = y/q \qquad E(R^2 I_{R\geq k}) = ky/q + 1$$

So $E(X|R\geq k) = \rho y/q$ $\qquad E(X^2|R\geq k) = 1 - \rho^2 + \rho^2(\frac{ky}{q} + 1) = 1 + \rho^2 \frac{ky}{q}$

and $\text{var}(X|Z\geq k) = 1 + \rho^2 ky/q + \rho^2 y^2/q^2$

$$= 1 + \rho^2 y/q\,(y/q - k)$$

If we have two such populations, male and female, then

$$E(X_m|R\geq k_m) = \rho_m \, {}^{y_m}/_{q_m} \qquad E(X_f|R\geq k_f) = \rho_f \, {}^{y_f}/_{q_f}$$

$$E(X_m^2|R\geq k_m) = \rho_m^2 \, {}^{k_m y_m}/_{q_m} + 1 \qquad E(X_f^2|\geq k_f) = \rho_f^2 \frac{k_f y_f}{q_f} + 1$$

$$\text{var}\,(X_m^2|R\geq k_m) = 1 - \rho_m^2 \, {}^{y_m}/_{q_m} \left({}^{y_m}/_{q_m} - k_m\right)$$

$$\text{var}\,(X_f^2|R\geq k_f) = 1 - \rho_f^2 \, {}^{y_f}/_{q_f} \left({}^{y_f}/_{q_f} - k_f\right)$$

If we pool the male and female population then

$$E(X_{m+f}) = \frac{q_m \bar{X}_m + q_f \bar{X}_f}{q_m + q_f} = \frac{\rho_m y_m + \rho_f y_f}{q_m + q_f}$$

$$E(X^2_{m+f}) = \frac{q_m \left(\rho_m^2 \, {}^{k_m y_m}/_{q_m} + 1\right) + q_f \left(\rho^2 \, {}^{k_f y_f}/_{q_f} + 1\right)}{q_m + q_f}$$

$$= 1 + \frac{\rho_m^2 k_m y_m + \rho_f^2 y_f y_f}{q_m + q_f}$$

$$\text{var}\,(X_{m+f}) = E(X^2_{m+f}) - E(X_{m+f})$$

REVISED
October 1983
EDITION

APPENDIX 5.1

ASAT GENERAL INFORMATION

INTERVIEW SCHEDULE

Name _____

School _____

Sex _____ Tape No. _____

Guarantee

We wish to reassure you that your replies to any questions in this interview will be treated as absolutely confidential.

At no stage will they be discussed with any of the teachers at your school.

There are no right or wrong answers to any question that may be put to you.

Counter No. [ ]

A	Do students feel that preparation/coaching make a difference to their mental state and/or performance?

A1 What preparation did you do for ASAT at school?

- the trial as an exam
- the trial but not as an exam
- multiple choice items like those on ASAT
- have you had much experience with multiple choice items?

A2 Did you do any study for ASAT at home?

- what did you do?
- was it useful?

A3 What advice were you given about ASAT?

- advice from parents, teachers, sisters/brothers
- did it include a strategy for attacking ASAT questions?
- did anyone suggest it was best to answer every question - even if it meant just guessing?

A4 Do you think you were prepared for ASAT?

If NO - did you have the opportunity to prepare?
- did that worry you?

If YES - did it make you feel more nervous or more confident?

All - was one trial paper enough?
- was the information bulletin sufficient?
- would extra practice at multiple choice items have helped?
- did you know what to expect when you went in to sit ASAT?

A5 Was ASAT like you expected it to be?

- harder
- easier
- totally different, how?
- had you been sufficiently prepared by the bulletins and the trial paper?

Counter No. [ ]

B Do students consider that boys have an advantage in ASAT because of
- greater willingness to engage in risk taking behaviour.
- item format favours boys.

B1 What was your overall strategy for doing the ASAT questions?

- did you go through from start to finish?
- did you jump those that were too difficult?
- did you leave any out because you couldn't do them?
- did you leave any out because you didn't get to them?
- did you fill in all the boxes by guessing? Why not?

B2 Did you have any concern about the multiple choice format - before?
- after?

- do you think it affected your performance?

B3 Were there any units that you particularly liked or disliked?

B4 Did you find yourself guessing more/less than you expected to have to?

Counter No. [ ]

C Do girls approach ASAT with greater apprehension?

C1 What concerned you most about ASAT before you went into the exam room?

- was that concern justified?

C2 Were you worried about the formality of the ASAT exam?

C3 Did your parents have attitudes that may have affected your performance?

- what about your teachers?

C4 Were you nervous - before ASAT?
- during ASAT?

- did this affect your performance?

C5 How important do you feel ASAT is?

- did this create extra pressure?

C6 Were you aware that ASAT was receiving a lot of comment in the press about sex bias?

- how did you react to this?
- did it affect your performance?
- did it create extra apprehension?

Counter No. [ ]

D Do students consider that those students with experience in mathematics/ science subjects have an advantage?

D1 Do you think students who study particular subjects have an advantage on ASAT?

- why?
- which subjects?

D2 What were your expectations of the amount of mathematics/science knowledge you would require for ASAT?

- were those expectations correct?

D3 Which subjects do you do?

D4-D8 Humanities students only

D4 Do you feel you would have done better if you had done more mathematics/science?

D5 Were you worried about any particular types of items?

- did you find that you had worried unnecessarily?

D6 How did you find the mathematics/science items?

- easy
- hard
- challenging
- boring

D7 What about the humanities items?

D8 Which unit did you find the

- easiest?
- hardest?

D9-D13 Maths/Science students only

137

D9 Do you feel you would have done better if you had done more humanities subjects?

D10 Were you worried about any particular types of items?

- did you find that you had worried unnecessarily?

D11 How were the mathematics/science items?

- hard
- easy
- boring
- challenging

D12 What about the humanities items?

D13 Which unit did you find the

- easiest?
- hardest?

Counter No. [ ]

E	Do differences exist between the sexes in their perceptions of whether or not ASAT is a valid indicator of scholastic aptitude?

E1 What is your opinion about the use of ASAT?

- do you know how it is used?

E2 What aspects of your 'personal standard of education' did ASAT.... (this generally will need explanation so agree on definition with other interviewer)

- test best?
- ignore?

E3 Do you think it can accurately measure your suitability for

- university entrance?
- other tertiary positions?
- a job?

E4 What did you see ASAT as?

- an IQ test
- something else

E5 What type of subjects did you expect ASAT to cover?

- traditional core?
- what did it actually cover?

E6 Did you feel you may not have had enough time to do it to the best of your ability?

E7 Compared with your results in other subjects how do you regard your results on ASAT? (Needs to be made clear that I am not asking how well you think you went but how important you regard them).

Counter No. [ ]

F Are girls more likely to be disadvantaged for physical and health reasons?

At this point the interviewer should be prepared to accept an answer of YES or NO. Do not probe for further information unless the students volunteer it.

F1 Did you have any physical or health problem which affected you on the day of testing?

- did it affect how you performed?

F2 Do you know of other students who may have been?

- how did test taking affect them?

Thank the student for his/her time and wind up the interview.

A.S.A.T. General Information Questionnaire

This questionnaire contains several questions about yourself and your feelings and reactions to A.S.A.T.

If you have trouble answering any questions, you may ask your teacher for help.

1. Name (please print): _____ _____
   Last name First name

2. Sex (tick one): ☐ Male 1 ☐ Female 2

3. Date of birth: _____ _____ 196_____
   Day Month Year

4. Present school: _____

In the items 5 - 26 tick only one box unless otherwise stated.

5. What kind of school course are you taking this year?

   Academic - mainly academic subjects leading to further education at a university or other tertiary institution. ☐ 1

   Vocational - mainly subjects to prepare for an occupation when you leave school. ☐ 2

6. Which subjects are you currently studying? (Tick more than one box)

   Art or Music ☐ 01
   Commerce, Economics or Business Studies ☐
   English ☐
   Geography, Psychology or Social Science ☐
   History ☐ 05
   Languages (apart from English) ☐
   Mathematics ☐

Physical Science (Physics, Chemistry) ☐

Biological Science (Biology, Environmental or Agricultural Science etc.) ☐

Technical Subjects (Metalwork, Woodwork, Typing etc.) ☐10

Other (please specify) ☐

7. In which year did you last study Mathematics?

1983 ☐1

1982 ☐

1981 ☐

1980 ☐

Earlier ☐5

8. In which year did you last study a foreign language?

1983 ☐1

1982 ☐

1981 ☐

1980 ☐

Earlier ☐5

Never ☐

9. In your present thoughts and plans, how much education do you intend to have yourself? (tick the highest amount you plan to have)

Year 12 ☐1

Post-secondary diploma or certificate course ☐2

University or college degree course ☐3

Other (say what) ☐4

10. Were there any formal preparation sessions for ASAT at school?

Yes ☐ 1

No ☐ 2

If yes how many did you attend? _____

11. Did you spend time preparing for ASAT at home?

Yes ☐ 1

No ☐ 2

If yes, approximately how many hours? _____

12. Please indicate the number of hours you had mathematics in 1982 and 1983.

	1982	1983
None	☐ 0	☐ 0
Less than 1 hour	☐ 1	☐ 1
between 1 hour and 1 hour 59 minutes	☐	☐
between 2 hours and 2 hours 59 minutes	☐	☐
between 3 hours and 3 hours 59 minutes	☐	☐
between 4 hours and 4 hours 59 minutes	☐ 5	☐ 5
between 5 hours and 5 hours 59 minutes	☐	☐
between 6 hours and 6 hours 59 minutes	☐	☐
between 7 hours and 7 hours 59 minutes	☐	☐
8 hours or more	☐ 9	☐ 9

13. In what country were each of you, your father and your mother born?
(Put one tick in each column)

	You	Father	Mother
Australia	☐ 1	☐ 1	☐ 1
Great Britian, Ireland, or New Zealand	☐	☐	☐
Italy	☐	☐	☐
Greece	☐	☐	☐
Yugoslavia	☐ 5	☐ 5	☐ 5
Elsewhere in Europe	☐	☐	☐
South East Asia	☐	☐	☐
Elsewhere in Asia	☐	☐	☐
Other	☐ 9	☐ 9	☐ 9

14. How many years have you lived in Australia? _____ years

15. Is English the only language used in your home?

Yes ☐ 1 No ☐ 2

If No

(a) Is English the language most commonly used in your home?

Yes ☐ 1 No ☐ 2

(b) What language apart from English is used? _____

16. How much education have your father and mother had?
(Put one tick in each column)

	Father	Mother
Primary school only	☐1	☐1
Some secondary school	☐	☐
Finished secondary school	☐	☐
Further training (not degree or diploma)	☐	☐
Tertiary (university, college degree or diploma)	☐5	☐5
Don't know	☐	☐

17. What is the present or last main occupation of your father or guardian?
(Name the occupation and describe what he does.)

Occupation: _____

What he does: _____

18. What is the present or last main occupation of your mother?
(Name the occupation and describe what she does.)

Occupation: _____

What she does: _____

The next eight questions relate to subject areas which you might no longer be studying. If so, answer them according to the last occasion on which you did so.

19. Compared to the marks I get in the other subjects I take, the marks I get in science subjects are

better	☐1
about the same	☐
worse	☐

20. Compared to the marks I get in the other subjects I take, the marks I get in mathematics are

- better ☐ 1
- about the same ☐
- worse ☐

21. Compared to the marks I get in other subjects, the marks I get in humanities subjects such as history or languages are

- better ☐ 1
- about the same ☐
- worse ☐

22. Compared to the marks I get in other subjects, the marks I get in social science subjects such as economics or geography are

- better ☐ 1
- about the same ☐
- worse ☐

23. I like science subjects

- more than other subjects ☐ 1
- about the same as other subjects ☐
- less than most other subjects ☐

24. I like mathematics subjects

- more than other subjects ☐ 1
- about the same as other subjects ☐
- less than most other subjects ☐

25. I like humanities subjects such as history or languages

more than other subjects	☐ 1
about the same as other subjects	☐
less than most other subjects	☐

26. I like social science subjects such as economics or geography

more than other subjects	☐ 1
about the same as other subjects	☐
less than most other subjects	☐

Please give your reaction to each of the following statements by ticking the box next to one of the categories.

SA if you strongly agree with the statement

A if you agree with the statement

D if you disagree with the statement

SD if you strongly disagree with the statement

U if you are uncertain or you don't know

		Strongly agree	Agree	Dis-agree	Strongly disagree	Un-certain
27.	Sitting for the A.S.A.T. was boring	☐ SA	☐ A	☐ D	☐ SD	☐ U
28.	I would rather have someone give me the answer to a difficult A.S.A.T. type question than have to find it out myself.	☐ SA	☐ A	☐ D	☐ SD	☐ U
29.	Even though I felt prepared, the A.S.A.T. questions seemed unusually hard for me.	☐ SA	☐ A	☐ D	☐ SD	☐ U
30.	Females certainly are logical enough to do well on A.S.A.T.	☐ SA	☐ A	☐ D	☐ SD	☐ U
31.	It would make me happy to be recognized as a student with high scholastic aptitude.	☐ SA	☐ A	☐ D	☐ SD	☐ U
32.	I generally felt confident about attempting A.S.A.T.	☐ SA	☐ A	☐ D	☐ SD	☐ U
33.	I felt that I could score well on A.S.A.T.	☐ SA	☐ A	☐ D	☐ SD	☐ U
34.	Females perform as well as males on A.S.A.T.	☐ SA	☐ A	☐ D	☐ SD	☐ U
35.	The challenge of A.S.A.T. type questions does not appeal to me.	☐ SA	☐ A	☐ D	☐ SD	☐ U

	Strongly agree	Agree	Dis-agree	Strongly disagree	Un-certain
36. It's hard to believe a female could do very well on A.S.A.T.	☐ SA	☐ A	☐ D	☐ SD	☐ U
37. I was no good at A.S.A.T.	☐ SA	☐ A	☐ D	☐ SD	☐ U
38. Compared with most of the students in my class, I probably went well on A.S.A.T.	☐ SA	☐ A	☐ D	☐ SD	☐ U
39. If I got a high score on A.S.A.T. I would only tell my parents.	☐ SA	☐ A	☐ D	☐ SD	☐ U
40. When an A.S.A.T. question arose that was difficult, I tried hard to give the best answer.	☐ SA	☐ A	☐ D	☐ SD	☐ U
41. Both males and females need preparation sessions for A.S.A.T.	☐ SA	☐ A	☐ D	☐ SD	☐ U
42. If I went well on A.S.A.T. I would try to keep it from my friends.	☐ SA	☐ A	☐ D	☐ SD	☐ U
43. The A.S.A.T. questions were enjoyable and stimulating to me.	☐ SA	☐ A	☐ D	☐ SD	☐ U
44. Males are naturally better than females on A.S.A.T. type questions.	☐ SA	☐ A	☐ D	☐ SD	☐ U
45. I'd be proud to be an outstanding student on A.S.A.T.	☐ SA	☐ A	☐ D	☐ SD	☐ U
46. Working out answers to A.S.A.T. questions did not interest me.	☐ SA	☐ A	☐ D	☐ SD	☐ U
47. I would expect a female who went well on A.S.A.T. to be rather unfeminine.	☐ SA	☐ A	☐ D	☐ SD	☐ U
48. I am not the type to do well on A.S.A.T.	☐ SA	☐ A	☐ D	☐ SD	☐ U

	Strongly agree	Agree	Dis-agree	Strongly disagree	Un-certain
49. I would be happy to get high scores on both the verbal and quantitative (mathematics / science) sections of A.S.A.T.	☐ SA	☐ A	☐ D	☐ SD	☐ U
50. I like doing A.S.A.T. type questions.	☐ SA	☐ A	☐ D	☐ SD	☐ U
51. People would think I was currying favour with the teachers if I tried to do really well on A.S.A.T.	☐ SA	☐ A	☐ D	☐ SD	☐ U
52. Females can do just as well as males on A.S.A.T.	☐ SA	☐ A	☐ D	☐ SD	☐ U
53. Most tests I handle well, but I messed up A.S.A.T.	☐ SA	☐ A	☐ D	☐ SD	☐ U
54. I wouldn't like people to think I was very good at A.S.A.T.	☐ SA	☐ A	☐ D	☐ SD	☐ U

55. Please comment here on any aspect of A.S.A.T. that is of interest or concern to you.

APPENDIX 6.2

FACTOR ANALYSIS OF ITEMS

Item No.	Scale 1						Scale 2						Scale 3						Scale 4					
	Males			Females			Males			Females			Males			Females			Males			Females		
	ACT	Qld	WA	ACT	Qld	WA	ACT	Qld	WA	ACT	Qld	WA	ACT	Qld	WA	ACT	Qld	WA	ACT	Qld	WA	ACT	Qld	WA
3	67	59	73	67	66	68																		
6	72	55	62	72	73	60						45												
7	74	66	70	74	77	65																		
11	74	72	72	79	71	81																		
12	70	58	64	70	67	73																		
22	70	71	67	65	65	76			30	34														
27	63	68	70	67	69	80	36																	
1							73	64	75	62	66	69												
2							42	51	51	37	48	57												
9						34	79	66	76	70	73	69												
17				30			80	67	77	72	68	75												
20							71	63	80	78	78	72												
24							78	71	78	76	72	77												
4													49	52	61	60	35	52						
8													57	68	79	75	57	75						
10													74	78	74	58	48	67						
18													72	78	82	74	46	65						
21													70	63	67		32	63				41		
26													66	75	69	83	57	58					37	
5		33															68		69	66	37	60		77
13																				33	61	54	74	
14											41				42		31		42	70		40		34
16													36							42	74	52	74	30
19									33								47		76	70	44	67		78
23																	53		69	63	52	66		66
28		32											45					48		30	52	52	36	
15		-35	-52	-35		-50					39	151							49			30		
25													58		31			38		52	49	30	33	
Percentage variance	14	13	14	15	15	16	13	12	14	13	13	13	13	12	13	10	9	12	8	10	8	10	7	8

APPENDIX 8.1

TRADITIONAL ITEM STATISTICS FOR ASAT-L

Humanities

Item number	Facility		r_{pbi} with subtest		r_{pbi} with full test	
	Male	Female	Male	Female	Male	Female
1	68.7	66.2	41	37	33	29
2	64.6	61.5	41	39	33	30
3	52.2	50.4	42	36	32	25
4	52.1	54.1	35	26	24	16
5	65.6	63.9	31	32	23	25
6	55.3	56.0	33	34	24	21
7	70.2	74.9	37	33	29	23
8	54.6	55.3	33	33	25	27
9	60.5	59.7	34	41	23	32
20	52.0	54.3	29	32	20	17
21	65.7	69.4	47	45	40	37
22	63.5	63.8	28	26	19	18
23	32.1	32.4	14	11	03	01
24	38.7	36.2	24	22	17	13
25	55.4	61.5	23	21	11	12
26	51.9	55.9	34	36	23	26
27	67.2	69.7	45	37	34	25
28	56.3	59.8	47	52	36	41
29	52.8	61.8	47	42	37	32
30	48.0	43.5	33	26	22	18
31	45.7	50.3	32	32	21	22
81	48.6	49.9	30	22	21	14
82	75.6	74.4	40	40	31	26
23	60.4	61.8	45	41	34	33
24	65.4	68.1	48	47	41	36
85	63.5	66.9	48	50	40	40
26	47.0	44.4	48	50	44	43
87	55.3	57.9	45	42	36	36
88	54.6	57.2	36	41	23	33
89	60.1	68.5	49	46	40	35

Social Science

Item number	Facility		r_{pbi} with subtest		r_{pbi} with full test	
	Male	Female	Male	Female	Male	Female
43	59.7	58.4	45	24	13	11
44	38.8	42.7	44	29	21	18
45	70.7	68.6	43	31	27	25
46	44.5	32.6	42	34	24	29
47	53.7	54.5	41	44	31	35
48	62.0	57.5	37	40	30	30
49	51.3	44.0	39	24	17	14
50	58.0	48.8	37	32	26	18
59	71.9	64.0	40	41	30	36
60	58.1	57.1	40	50	40	43
61	60.0	52.6	36	43	37	34
62	46.0	44.6	38	48	36	39
63	61.9	65.6	38	37	33	30
64	66.1	65.9	37	26	19	13
65	55.0	42.1	32	34	33	27
66	84.6	83.8	28	38	35	34
67	50.3	47.9	28	23	11	10
68	71.4	68.2	29	34	25	26
69	67.0	64.2	26	35	30	25
70	47.8	50.4	23	29	33	20

Science

Item number	Facility		r_{pbi} with subtest		r_{pbi} with full test	
	Male	Female	Male	Female	Male	Female
10	92.3	78.3	57	53	34	47
11	88.2	72.1	48	53	40	46
12	88.9	83.8	47	35	22	27
13	83.2	64.5	45	51	38	43
14	60.9	40.0	41	44	39	39
15	85.8	79.4	47	36	21	30
32	93.8	94.8	47	24	25	21
33	48.8	33.0	47	43	42	39
34	56.1	43.1	46	34	36	25
35	61.2	40.6	42	48	49	40
36	58.8	45.4	42	50	40	42
37	47.0	41.0	43	41	33	34
51	87.2	88.6	30	33	31	33
52	73.9	72.7	33	30	30	23
53	80.7	72.2	41	40	32	33
54	58.4	54.6	34	37	36	33
55	57.0	48.7	36	34	35	28
75	89.1	89.7	43	34	29	29
76	59.0	51.0	27	50	44	48
77	55.9	51.0	41	28	26	21
78	70.3	64.6	47	51	42	44
79	75.8	69.1	34	43	36	37
80	80.5	79.0	40	44	35	36
94	83.4	81.1	35	39	35	36
95	64.9	62.3	39	36	33	29
96	78.3	78.0	38	40	27	32
97	75.4	73.1	36	38	32	31
92	65.7	57.9	34	29	28	24
99	66.1	64.2	35	39	29	32
100	36.0	32.2	27	27	25	22

Mathematics

Item number	Facility		r_{pbi} with subtest		r_{pbi} with full test	
	Male	Female	Male	Female	Male	Female
16	52.0	50.6	34	40	28	34
17	49.0	44.0	34	32	29	27
18	77.6	73.3	43	46	36	40
19	70.9	70.2	42	41	30	30
38	68.1	58.8	49	46	36	32
39	81.9	76.2	43	44	32	35
40	62.2	55.3	44	36	29	23
41	60.0	52.0	40	36	31	23
42	48.6	38.5	43	30	34	22
56	58.0	52.2	30	30	21	21
57	68.0	59.9	32	32	23	20
58	82.3	78.6	31	34	26	28
71	54.7	44.4	50	49	37	35
72	66.2	55.9	50	47	38	34
73	75.9	66.9	33	34	20	22
74	83.2	74.1	35	37	27	31
90	67.2	51.0	51	56	39	45
91	71.9	72.3	24	26	14	16
92	72.5	58.1	51	52	39	42
93	47.0	36.9	47	38	41	30

APPENDIX 8.2

RESULTS OF FACTOR ANALYSES OF ITEMS AND VARIMAX ROTATION WITH TWO FACTORS[a]

Item	Males		Females		Both sexes		Scale[b]	Quantitative or verbal scale[c]
	I	II	I	II	I	II		
1		39		35		36	H	V
2		41		34		37	H	V
3		41		32		36	H	V
4		34		21		28	H	V
5		25		25		26	H	V
6		27		33		30	H	V
7		36		33		35	H	V
8		29		30		30	H	V
9		31		39		35	H	V
10	34	12	46	21	48	16	S	Q
11	39	23	50	20	50	17	S	Q
12	25		33		31		S	Q
13	40	17	50		50		S	Q
14	47		44		50		S	Q
15	16	16	30	15	26		S	Q
16	29		34	19	30	16	M	Q
17	31		25	17	28		M	Q
18	45		43	18	43	15	M	Q
19	41		35		36		M	Q
20		23		33		28	H	V
21	18	43	16	43	16	44	H	V
22		23		24		23	H	V
23[d]							H	omit
24		19				15	H	V
25		18		17		19	H	V
26		30		33		32	H	V
27		43		35		39	H	V
28		46	15	50		48	H	V
29		43		40		44	H	V
30		30		21		24	H	V
31		29		27		29	H	V
32	25		17	15	20	16	S	Q
33	47	16	46		49		S	Q
34	44		35		41		S	Q
35	57	17	48		55		S	Q
36	46		47	15	48		S	Q
37	37		39		38		S	Q
38	50		46		49		M	Q
39	46		44		45		M	Q
40	42		32		37		M	Q
41	37		32		35		M	Q
42	43		28		37		M	Q

(Continued)

Appendix 8.2 (Continued)

Item	Males		Females		Both sexes		Scale[b]	Quantitative or verbal scale[c]
	I	II	I	II	I	II		
43	19				15		SS	Q
44	22				15	15	SS	V
45	23	17	31		27		SS	Q
46	24		35		33		SS	Q
47	18	30	29	24	23	28	SS	V
48	31		28	18	30	16	SS	Q
49	20				19		SS	Q
50	31		19		27		SS	Q
51	16	34	20	32	17	33	S	V
52	26	20	23		24	16	S	Q
53	33	15	28	23	33	18	S	Q
54	29	27	23	29	26	28	S	V
55	32	20	22	22	29	20	S	Q
56	23		23		24		M	Q
57	25		27		28		M	Q
58	26		31		29		M	Q
59	17	30	19	38	21	32	SS	V
60	19	44	20	48	20	45	SS	V
61	20	38		39	20	37	SS	V
62		41	19	41	18	41	SS	V
63		41	17	32		37	SS	V
64		29		23		25	SS	V
65	34	16	28		34		SS	Q
66		40		43		41	SS	V
67				17			SS	V
68	29		27		28		SS	Q
69	19	28	19	21	19	24	SS	V
70	18	32		23		29	SS	V
71	47		45		47		M	Q
72	47		43		46		M	Q
73	30		30		31		M	Q
74	34		34		36		M	Q
75	16	29	19	28	17	29	S	V
76	32	36	36	39	36	36	S	V
77	23	17	18		21	15	S	Q
78	32	34	42	26	37	30	S	Q
79	35	20	35	22	36	21	S	Q
80	27	29	32	25	29	27	S	Q
81		24				19	H	V
82		41		41		41	H	V
83		44	17	36		40	H	V
84	15	48		46		47	H	V
85	15	46		49		48	H	V
86	26	41	21	46	24	44	H	V
87		42	17	40		42	H	V
88		31	18	34		34	H	V
89		48		44		48	H	V

(Continued)

Appendix 8.2 (Continued)

Item	Males		Females		Both sexes		Scale[b]	Quantitative or verbal scale[c]
	I	II	I	II	I	II		
90	54		54		56		M	Q
91	17		20		17		M	Q
92	52		51		53		M	Q
93	46	15	38		43		M	Q
94	34	20	31	23	32	22	S	Q
95	28	23	26	19	26	22	S	Q
96	28		35		30	15	S	Q
97	32	17	29	19	30	19	S	Q
98	27	16	26		28		S	Q
99	25	19	33	15	29	18	S	Q
100	23		23		23		S	Q

[a] Only loadings greater than 0.15 have been listed, decimal points have been omitted.

[b] H - Humanities; S - Science; M - Mathematics; SS - Social Science

[c] V - Verbal; Q - Quantitative

[d] Item 23 was omitted from all analyses because it had unsatisfactory item statistics.

APPENDIX 8.3

RESULTS OF FACTOR ANALYSES OF UNITS AND VARIMAX ROTATION WITH TWO FACTORS[a]

Unit	Scale[b]	Males		Females		Both sexes	
		I	II	I	II	I	II
1	H		76		74		75
2	S	55	32	62	35	64	
3	M	59		55	31	56	
4	H		77		74		76
5	S	67		66		69	
6	N	68		65		67	
7	SS	48	31	50		50	
8	S	46	44	34	50	42	46
9	M	41		44		44	
10	SS	34	69		71	33	69
11	M	62		60		62	
12	S	42	54	46	52	45	52
13	H		77		75		77
14	M	69		67		69	
15	S	50	38	53	35	51	37

a Only loadings greater than 0.30 have been listed and decimal points have been omitted.

b H - Humanities; S - Science; M - Mathematics; SS - Social Science

APPENDIX 8.4

RASCH ANALYSES ITEM STATISTICS FOR ASAT-L

Humanities Items

	Item Difficulty[a]				Item between Group Fit			
	Full test		Subtest		Full test		Subtest	
Item	Male	Female	Male	Female	Male	Female	Male	Female
1	-0.20	-0.28	-0.57	-0.37	0.69	-0.88	1.13	0.95
2	0.00	-0.01	-0.36	-0.15	1.47	-0.80	0.85	0.14
3	0.58	0.46	0.22	0.38	-1.46	1.40	1.02	-0.29
4	0.58	0.27	0.22	0.19	1.91	4.41	-0.52	2.61
5	-0.05	-0.18	-0.41	-0.26	2.17	1.05	0.50	1.15
6	0.43	0.22	0.08	0.14	4.71	2.16	1.52	-1.24
7	-0.29	-0.75	-0.64	-0.83	-0.44	-1.03	1.01	0.63
8	0.46	0.23	0.11	0.15	2.[illegible]	-0.64	1.41	0.57
9	0.19	0.01	-0.17	-0.07	1.98	0.17	-0.39	0.99
20	0.58	0.28	0.23	0.20	2.88	5.05	2.17	0.51
21	-0.05	-0.42	-0.41	-0.51	3.26	3.18	3.32	3.95
22	0.02	-0.16	-0.30	-0.24	2.31	4.32	2.66	2.47
23	1.47	1.27	1.16	1.21	7.92	9.01	7.07	7.96
24	1.17	1.09	0.84	1.02	4.16	6.20	3.33	5.30
25	0.42	-0.06	0.07	-0.13	6.43	5.09	4.78	4.63
26	0.58	0.19	0.23	0.11	2.40	1.04	0.25	-0.46
27	-0.13	-0.45	-0.49	-0.54	0.33	0.07	2.67	0.18
28	0.40	0.02	0.03	-0.06	0.92	3.21	3.56	5.39
29	0.55	-0.07	0.19	-0.15	2.36	0.49	3.25	2.60
30	0.75	0.74	0.41	0.67	2.71	3.21	-0.39	3.84
31	0.88	0.44	0.52	0.37	3.57	2.77	0.49	1.62
81	0.73	0.47	0.38	0.39	3.12	4.99	1.56	4.42
82	-0.58	-0.70	-0.95	-0.79	2.11	-2.13	2.61	2.81
83	0.20	-0.09	-0.16	-0.16	-0.93	0.58	1.97	1.50
84	-0.04	-0.37	-0.40	-0.46	3.18	1.68	3.70	3.40
85	0.06	-0.30	-0.31	-0.39	3.71	3.88	4.40	5.66
86	0.18	0.72	0.45	0.64	4.50	4.76	3.55	5.22
87	0.44	0.11	0.08	0.03	1.21	0.50	2.12	2.08
88	0.47	0.14	0.11	0.07	3.01	1.33	-0.09	1.99
89	0.22	-0.39	-0.15	-0.48	4.79	1.76	4.37	3.47

a Item difficulties expressed in logits.

Social Science Items

Item	Item Difficulty[a]				Item between Group Fit			
	Full test		Subtest		Full test		Subtest	
	Male	Female	Male	Female	Male	Female	Male	Female
43	0.23	0.08	0.00	-0.11	6.50	6.69	3.10	2.60
44	1.17	0.79	0.94	0.59	2.14	3.65	1.30	0.98
45	-0.31	-0.39	-0.55	-0.58	-1.84	-0.16	0.31	0.92
46	0.93	1.29	0.67	1.08	1.83	-0.34	-0.25	-0.84
47	0.49	0.25	0.26	0.05	1.08	0.82	2.42	2.77
48	0.13	0.12	-0.11	-0.07	0.46	0.44	2.04	1.87
49	0.61	0.73	0.38	0.53	4.51	5.28	1.16	3.27
50	0.32	0.54	0.07	0.34	1.50	4.25	1.44	0.99
59	-0.37	-0.18	-0.62	-0.37	0.70	2.18	1.24	3.19
60	0.31	0.14	0.06	-0.05	3.01	4.47	2.69	5.50
61	0.24	0.35	-0.01	0.14	1.08	0.35	2.52	1.42
62	0.86	0.70	0.60	0.49	1.09	2.11	1.69	4.67
63	0.13	-0.25	-0.11	-0.44	-0.49	0.30	1.77	1.18
64	-0.07	-0.25	-0.31	-0.43	3.39	4.91	2.53	0.67
65	0.45	0.83	0.20	0.63	1.59	0.88	1.30	1.38
66	-1.20	-1.32	-1.45	-1.49	3.04	3.21	4.02	4.21
67	0.65	0.55	0.42	0.36	6.92	6.90	3.08	3.83
68	-0.35	-0.37	-0.59	-0.56	0.97	-1.08	-1.05	-0.69
69	-0.12	-0.18	-0.36	-0.37	0.40	0.56	1.65	-0.32
70	0.77	0.45	0.52	0.25	2.03	3.21	-0.03	0.61

a Item difficulties expressed in logits.

Science Items

Item	Item Difficulty[a]				Item between Group Fit			
	Full test		Subtest		Full test		Subtest	
	Male	Female	Male	Female	Male	Female	Male	Female
10	-2.03	-0.95	-1.67	-0.75	4.41	6.98	3.01	6.97
11	-1.53	-0.58	-1.16	-0.36	5.14	6.12	4.28	5.21
12	-1.61	-1.31	-1.23	-1.13	-0.40	1.05	0.16	1.55
13	-1.09	-0.19	-0.69	0.06	3.47	4.32	3.01	3.41
14	0.18	0.94	0.66	1.27	1.77	2.33	1.23	1.57
15	-1.30	-1.01	-0.91	-0.81	1.26	0.00	4.46	-1.60
32	-2.26	-2.61	-1.90	-2.45	2.22	1.52	0.57	-0.41
33	0.74	1.26	1.26	1.63	2.93	2.69	1.43	1.96
34	0.40	0.79	0.90	1.11	2.12	2.41	0.56	2.92
35	0.17	0.88	0.64	1.21	5.73	3.42	5.39	2.21
36	0.28	0.67	0.77	0.98	1.50	3.70	1.98	3.87
37	0.81	0.88	1.35	1.21	-0.53	1.15	-0.25	0.56
51	-1.43	-1.75	-1.04	-1.57	2.17	3.99	-0.35	1.42
52	-0.48	-0.62	-0.05	-0.40	-1.04	0.54	0.78	1.87
53	-0.91	-0.58	-0.50	-0.36	0.51	0.97	0.28	-1.02
54	0.30	0.25	0.79	0.54	-0.05	1.17	-0.12	3.20
55	0.39	0.52	0.86	0.83	1.85	0.29	0.96	2.25
75	-1.62	-1.88	-1.25	-1.71	0.89	2.70	-0.51	1.87
76	0.29	0.42	0.76	0.71	3.71	5.98	1.78	2.19
77	0.41	0.42	0.91	0.71	0.48	2.94	1.44	5.77
78	-0.29	-0.21	0.16	0.04	3.98	4.65	2.03	3.25
79	-0.59	-0.43	-0.17	-0.20	1.49	2.64	1.47	1.31
80	-0.90	-0.99	-0.49	-0.79	2.34	2.78	1.70	2.98
94	-1.11	-1.12	-0.71	-0.92	2.59	2.85	2.24	1.02
95	-0.02	-0.10	0.46	0.16	0.21	0.78	1.49	1.37
96	-0.75	-0.94	-0.33	-0.74	-0.09	1.15	0.18	0.48
97	-0.58	-0.64	-0.14	-0.42	0.10	0.15	0.57	-0.09
98	-0.05	0.12	0.42	0.39	-0.47	1.46	1.25	4.64
99	-0.07	-0.18	0.39	0.07	-0.28	1.41	1.26	-0.70
100	1.34	1.31	1.91	1.69	1.18	2.76	4.88	4.99

a Item difficulties expressed in logits.

Mathematics Items

Item	Item Difficulty[a]				Item between Group Fit			
	Full test		Subtest		Full test		Subtest	
	Male	Female	Male	Female	Male	Female	Male	Female
16	0.59	0.42	0.73	0.37	2.03	-1.80	1.83	-0.07
17	0.73	0.74	0.88	0.70	1.36	0.94	2.87	2.07
18	-0.70	-0.65	-0.62	-0.73	-1.83	-3.23	1.77	3.65
19	-0.32	-0.47	-0.22	-0.54	-0.39	-0.79	0.23	1.49
38	-0.17	0.07	-0.08	0.00	1.71	-1.02	2.63	2.35
39	-0.99	-0.82	-0.94	-0.92	1.25	2.11	2.64	3.53
40	0.81	0.24	0.23	0.18	-0.26	1.40	-0.53	0.36
41	0.22	0.40	0.34	0.35	0.95	1.82	0.97	1.13
42	0.75	1.01	0.91	0.98	0.51	3.43	1.02	3.34
56	0.30	0.36	0.45	0.30	2.24	2.20	4.31	3.82
57	-0.17	0.01	-0.07	-0.06	2.09	3.39	2.53	1.78
58	-1.02	-0.96	-0.96	-1.07	0.36	-0.40	1.82	1.53
71	0.47	0.72	0.60	0.68	1.39	0.56	2.66	2.87
72	-0.08	0.20	0.03	0.14	2.06	0.62	2.73	3.14
73	-0.60	-0.32	-0.52	-0.39	1.96	1.20	0.93	-0.41
74	-1.17	-0.69	-1.11	-0.79	0.69	2.72	-0.50	-0.25
90	-0.12	0.43	-0.03	0.37	1.85	4.53	3.75	6.12
91	-0.37	-0.57	-0.28	-0.65	4.54	4.05	5.65	3.77
92	-0.40	0.10	-0.32	0.03	2.45	3.77	4.42	4.40
93	0.82	1.08	0.98	1.06	2.43	4.09	0.88	2.48

a Item difficulties expressed in logits.

The Australian Scholastic Aptitude Test (ASAT) was first used as a moderating device in the Australian Capital Territory in 1977. Differences have been observed in the average performance of males and females on the test. The study reported in this monograph investigated the nature and origin of the score differences recorded for ASAT. The research focused on five issues: retention rate differences, attitude differences, preparation differences, item bias, effect of differential course selection. The findings indicated that sex had no significant direct effect on ASAT scores. The observed differences in male and female ASAT scores were related to English ability, experience in mathematics, and confidence in success. The different patterns for staying longer at school could explain much of the observed variations between males and females.

Australian Council for Educational Research
ISBN 0 85563 376 X